

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
10 May 2002 (10.05.2002)

PCT

(10) International Publication Number  
**WO 02/36807 A2**

- (51) International Patent Classification<sup>7</sup>: **C12Q** Leah, R. [US/US]; 939 Rosette Court, Sunnyvale, CA 94086 (US).
- (21) International Application Number: **PCT/US01/46888**
- (22) International Filing Date: 29 October 2001 (29.10.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/243,802 27 October 2000 (27.10.2000) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:  
US 60/243,802 (CIP)  
Filed on 27 October 2000 (27.10.2000)
- (71) Applicant (for all designated States except US): **DI-ADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard, South San Francisco, CA 94080 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **SUN, Yongming** [US/US]; Apartment 260, 869 S. Winchester Boulevard, San Jose, CA 95128 (US). **RECIPON, Herve** [FR/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). **SALCEDA, Susana** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way, #14, San Jose, CA 95117 (US). **TURNER,**
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:  
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/36807 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO BREAST SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic breast cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast tissue, identifying breast tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered breast tissue for treatment and research.

## COMPOSITIONS AND METHODS RELATING TO BREAST SPECIFIC GENES AND PROTEINS

This application claims the benefit of priority from U.S. Provisional Application  
5 Serial No. 60/243,805 filed October 27, 2000, which is herein incorporated by reference  
in its entirety.

### FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and  
10 polypeptides present in normal and neoplastic breast cells, including fragments, variants  
and derivatives of the nucleic acids and polypeptides. The present invention also relates  
to antibodies to the polypeptides of the invention, as well as agonists and antagonists of  
the polypeptides of the invention. The invention also relates to compositions comprising  
the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists  
15 of the invention and methods for the use of these compositions. These uses include  
identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-  
cancerous disease states in breast tissue, identifying breast tissue and monitoring and  
identifying and/or designing agonists and antagonists of polypeptides of the invention.  
The uses also include gene therapy, production of transgenic animals and cells, and  
20 production of engineered breast tissue for treatment and research.

### BACKGROUND OF THE INVENTION

Excluding skin cancer, breast cancer, also called mammary tumor, is the most  
common cancer among women, accounting for a third of the cancers diagnosed in the  
United States. One in nine women will develop breast cancer in her lifetime and about  
25 192,000 new cases of breast cancer are diagnosed annually with about 42,000 deaths.  
Bevers, *Primary Prevention of Breast Cancer*, in BREAST CANCER, 20-54 (Kelly K Hunt  
et al., ed., 2001); Kochanek et al., 49 Nat'l. Vital Statistics Reports 1, 14 (2001).

In the treatment of breast cancer, there is considerable emphasis on detection and  
risk assessment because early and accurate staging of breast cancer has a significant  
30 impact on survival. For example, breast cancer detected at an early stage (stage T0,  
discussed below) has a five-year survival rate of 92%. Conversely, if the cancer is not

detected until a late stage (i.e., stage T4), the five-year survival rate is reduced to 13%.  
AJCC Cancer Staging Handbook pp. 164-65 (Irvin D. Fleming et al. eds., 5<sup>th</sup> ed. 1998).

Some detection techniques, such as mammography and biopsy, involve increased discomfort, expense, and/or radiation, and are only prescribed only to patients with an  
5 increased risk of breast cancer.

Current methods for predicting or detecting breast cancer risk are not optimal. One method for predicting the relative risk of breast cancer is by examining a patient's risk factors and pursuing aggressive diagnostic and treatment regiments for high risk patients. A patient's risk of breast cancer has been positively associated with increasing  
10 age, nulliparity, family history of breast cancer, personal history of breast cancer, early menarche, late menopause, late age of first full term pregnancy, prior proliferative breast disease, irradiation of the breast at an early age and a personal history of malignancy. Lifestyle factors such as fat consumption, alcohol consumption, education, and socioeconomic status have also been associated with an increased incidence of breast  
15 cancer although a direct cause and effect relationship has not been established. While these risk factors are statistically significant, their weak association with breast cancer limited their usefulness. Most women who develop breast cancer have none of the risk factors listed above, other than the risk that comes with growing older. NIH Publication No. 00-1556 (2000).

20 Current screening methods for detecting cancer, such as breast self exam, ultrasound, and mammography have drawbacks that reduce their effectiveness or prevent their widespread adoption. Breast self exams, while useful, are unreliable for the detection of breast cancer in the initial stages where the tumor is small and difficult to detect by palpitation. Ultrasound measurements require skilled operators at an increased  
25 expense. Mammography, while sensitive, is subject to over diagnosis in the detection of lesions that have questionable malignant potential. There is also the fear of the radiation used in mammography because prior chest radiation is a factor associated with an increase incidence of breast cancer.

At this time, there are no adequate methods of breast cancer prevention. The  
30 current methods of breast cancer prevention involve prophylactic mastectomy (mastectomy performed before cancer diagnosis) and chemoprevention (chemotherapy

before cancer diagnosis) which are drastic measures that limit their adoption even among women with increased risk of breast cancer. Bevers, *supra*.

A number of genetic markers have been associated with breast cancer. Examples of these markers include carcinoembryonic antigen (CEA) (Mughal et al., 249 JAMA 1881 (1983)) MUC-1 (Frische and Liu, 22 J. Clin. Ligand 320 (2000)), HER-2/neu (Haris et al., 15 Proc.Am.Soc.Clin.Oncology. A96 (1996)), uPA, PAI-1, LPA, LPC, RAK and BRCA (Esteva and Fritsche, *Serum and Tissue Markers for Breast Cancer*, in BREAST CANCER, 286-308 (2001)). These markers have problems with limited sensitivity, low correlation, and false negatives which limit their use for initial diagnosis. For example, while the BRCA1 gene mutation is useful as an indicator of an increased risk for breast cancer, it has limited use in cancer diagnosis because only 6.2 % of breast cancers are BRCA1 positive. Maloné et al., 279 JAMA 922 (1998). *See also*, Mewman et al., 279 JAMA 915 (1998) (correlation of only 3.3%).

Breast cancers are diagnosed into the appropriate stage categories recognizing that different treatments are more effective for different stages of cancer. Stage TX indicates that primary tumor cannot be assessed (i.e., tumor was removed or breast tissue was removed). Stage T0 is characterized by abnormalities such as hyperplasia but with no evidence of primary tumor. Stage Tis is characterized by carcinoma in situ, intraductal carcinoma, lobular carcinoma in situ, or Paget's disease of the nipple with no tumor. Stage T1 is characterized as having a tumor of 2 cm or less in the greatest dimension. Within stage T1, Tmic indicates microinvasion of 0.1 cm or less, T1a indicates a tumor of between 0.1 to 0.5 cm, T1b indicates a tumor of between 0.5 to 1 cm, and T1c indicates tumors of between 1 cm to 2 cm. Stage T2 is characterized by tumors from 2 cm to 5 cm in the greatest dimension. Tumors greater than 5 cm in size are classified as stage T4. Within stage T4, T4a indicates extension of the tumor to the chest wall, T4b indicates edema or ulceration of the skin of the breast or satellite skin nodules confined to the same breast, T4c indicates a combination of T4a and T4b, and T4d indicates inflammatory carcinoma. AJCC Cancer Staging Handbook pp. 159-70 (Irvin D. Fleming et al. eds., 5<sup>th</sup> ed. 1998). In addition to standard staging, breast tumors may be classified according to their estrogen receptor and progesterone receptor protein status. Fisher et al., 7 Breast Cancer Research and Treatment 147 (1986). Additional pathological status, such as HER2/neu status may also be useful. Thor et al., 90



J.Nat'l.Cancer Inst. 1346 (1998); Paik et al., 90 J.Nat'l.Cancer Inst. 1361 (1998); Hutchins et al., 17 Proc.Am.Soc.Clin.Oncology A2 (1998).; and Simpson et al., 18 J.Clin.Oncology 2059 (2000).

In addition to the staging of the primary tumor, breast cancer metastases to  
5 regional lymph nodes may be staged. Stage NX indicates that the lymph nodes cannot be assessed (e.g., previously removed). Stage N0 indicates no regional lymph node metastasis. Stage N1 indicates metastasis to movable ipsilateral axillary lymph nodes. Stage N2 indicates metastasis to ipsilateral axillary lymph nodes fixed to one another or to other structures. Stage N3 indicates metastasis to ipsilateral internal mammary lymph  
10 nodes. Id.

Stage determination has potential prognostic value and provides criteria for designing optimal therapy. Simpson et al., 18 J. Clin. Oncology 2059 (2000). Generally, pathological staging of breast cancer is preferable to clinical staging because the former gives a more accurate prognosis. However, clinical staging would be preferred if it were  
15 as accurate as pathological staging because it does not depend on an invasive procedure to obtain tissue for pathological evaluation. Staging of breast cancer would be improved by detecting new markers in cells, tissues, or bodily fluids which could differentiate between different stages of invasion. Progress in this field will allow more rapid and reliable method for treating breast cancer patients.

20 Treatment of breast cancer is generally decided after an accurate staging of the primary tumor. Primary treatment options include breast conserving therapy (lumpectomy, breast irradiation, and surgical staging of the axilla), and modified radical mastectomy. Additional treatments include chemotherapy, regional irradiation, and, in extreme cases, terminating estrogen production by ovarian ablation.

25 Until recently, the customary treatment for all breast cancer was mastectomy. Fonseca et al., 127 Annals of Internal Medicine 1013 (1997). However, recent data indicate that less radical procedures may be equally effective, in terms of survival, for early stage breast cancer. Fisher et al., 16 J. of Clinical Oncology 441 (1998). The treatment options for a patient with early stage breast cancer (i.e., stage Tis) may be  
30 breast-sparing surgery followed by localized radiation therapy at the breast. Alternatively, mastectomy optionally coupled with radiation or breast reconstruction may

be employed. These treatment methods are equally effective in the early stages of breast cancer.

Patients with stage I and stage II breast cancer require surgery with chemotherapy and/or hormonal therapy. Surgery is of limited use in Stage III and stage IV patients.

- 5 Thus, these patients are better candidates for chemotherapy and radiation therapy with surgery limited to biopsy to permit initial staging or subsequent restaging because cancer is rarely curative at this stage of the disease. AJCC Cancer Staging Handbook 84, ¶. 164-65 (Irvin D. Fleming et al. eds., 5<sup>th</sup> ed. 1998).

- 10 In an effort to provide more treatment options to patients, efforts are underway to define an earlier stage of breast cancer with low recurrence which could be treated with lumpectomy without postoperative radiation treatment. While a number of attempts have been made to classify early stage breast cancer, no consensus recommendation on postoperative radiation treatment has been obtained from these studies. Page et al., 75 Cancer 1219 (1995); Fisher et al., 75 Cancer 1223 (1995); Silverstein et al., 77 Cancer  
15 2267 (1996).

- As discussed above, each of the methods for diagnosing and staging breast cancer is limited by the technology employed. Accordingly, there is need for sensitive molecular and cellular markers for the detection of breast cancer. There is a need for molecular markers for the accurate staging, including clinical and pathological staging, of  
20 breast cancers to optimize treatment methods. Finally, there is a need for sensitive molecular and cellular markers to monitor the progress of cancer treatments, including markers that can detect recurrence of breast cancers following remission.

- Other objects, features, advantages and aspects of the present invention will become apparent to those of skill in the art from the following description. It should be  
25 understood, however, that the following description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only. Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following description and from reading the other parts of the present disclosure.

## SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto that may be used to identify, diagnose, monitor, stage, image and treat breast cancer and  
5 non-cancerous disease states in breast; identify and monitor breast tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered breast tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that  
10 are specific to breast cells and/or breast tissue. These breast specific nucleic acids (BSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the BSNA is genomic DNA, then the BSNA is a breast specific gene (BSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to  
15 breast. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 160 through 282. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 159. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence  
20 similarity to a nucleic acid molecule encoding a BSP, or that selectively hybridize or exhibit substantial sequence similarity to a BSNA, as well as allelic variants of a nucleic acid molecule encoding a BSP, and allelic variants of a BSNA. Nucleic acid molecules comprising a part of a nucleic acid sequence that encodes a BSP or that comprises a part of a nucleic acid sequence of a BSNA are also provided.

25 A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of a BSNA. In a preferred embodiment, the nucleic acid molecule comprises one or more expression control sequences controlling the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment  
30 of a BSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic

acid molecule encodes all or a fragment of a BSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of a BSNA.

Another object of the invention is to provide methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly  
5 produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is a BSP. The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutein), fusion protein, homologous protein or a polypeptide encoded by  
10 an allelic variant of a BSP.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention..

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

15 Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and  
20 treating breast cancer and non-cancerous disease states in breast. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring breast tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered breast tissue for treatment and research.

25 The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat breast cancer and non-cancerous disease states in breast. The invention provides methods of using the polypeptides of the invention to identify and/or monitor breast tissue, and to produce engineered breast tissue.

30 The agonists and antagonists of the instant invention may be used to treat breast cancer and non-cancerous disease states in breast and to produce engineered breast tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences.

## 5 DETAILED DESCRIPTION OF THE INVENTION

### Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular  
10 terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed  
15 according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.,* Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor  
20 Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4<sup>th</sup> Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of  
25 which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and  
30 techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in

the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

- 5           A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and
- 10   "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.
- 15           The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates,
- 20   phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially
- 25   duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.
- 30           A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may

comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a  
5 nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

10 A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an  
15 RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a  
20 portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized  
25 polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

30 A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid

molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a  
5 nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the  
10 disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other  
15 nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others.  
20 Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense  
25 oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of  
30 other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are



not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized  
5 oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

10 The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate,  
15 phosphorodiselenoate, phosphoroanilothioate, phosphoraniladate, phosphoroamidate, and the like. See e.g., LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); United States Patent No.  
20 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is  
25 referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of  
30 RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence.

In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum  
5 correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide  
10 sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson,  
15 *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with  
20 its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular  
25 sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology"  
30 interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point ( $T_m$ ) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the  $T_m$  for the specific DNA hybrid under a particular set of conditions. The  $T_m$  is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

The  $T_m$  for a particular DNA-DNA hybrid can be estimated by the formula:

$$T_m = 81.5^\circ\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G} + \text{C}) - 0.63 (\% \text{ formamide}) - (600/l)$$

where  $l$  is the length of the hybrid in base pairs.

The  $T_m$  for a particular RNA-RNA hybrid can be estimated by the formula:

5 
$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.35 (\% \text{ formamide}) - (820/l).$$

The  $T_m$  for a particular RNA-DNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.50 (\% \text{ formamide}) - (820/l).$$

10 In general, the  $T_m$  decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C  
15 would be subtracted from the calculated  $T_m$  of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

20 An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without  
25 formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a  
30 library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping

- the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (*e.g.* 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

- As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon degeneracy as permitted by the redundancy of the genetic code.

- Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula:  $T_m = 81.5^{\circ}\text{C} + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\text{fraction G+C}) - (600/N)$ , wherein N is change length and the  $[\text{Na}^+]$  is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the  $T_m$ ) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The

- various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1  $\mu$ g of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20  $\mu$ l of reaction buffer. For the
- 5 purpose of isolating DNA fragments for plasmid construction, typically 5 to 50  $\mu$ g of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are
- 10 ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.
- 15 The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.
- Genome-derived "single exon probes," are probes that comprise at least part of an
- 20 exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the
- 25 exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon
- 30 probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity

with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid molecule comprises the wild type nucleic acid sequence encoding a BSP or is a BSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR

reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. See, e.g., Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

10 The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate  
15 random mutations within the DNA.

The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

20 The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. See, e.g., Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815  
25 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. See, e.g., Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455  
30 (1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.



"Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

5           The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination,  
10   promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such  
15   control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

20           The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of  
25   vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and  
30   thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression

vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that  
5 serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding  
10 generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in  
15 its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

20 As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-  
25 occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a BSP encoded by a nucleic acid molecule of the instant  
30 invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally

associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be

5 "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single

10 species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample,

15 followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared

20 to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40

25 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such

30 modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid

derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, 5 proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of 10 substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as  $^{125}\text{I}$ ,  $^{32}\text{P}$ ,  $^{35}\text{S}$ , and  $^3\text{H}$ , ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, 15 stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences. 20 Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be 25 produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

30 The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial

identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with  
5 respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide of the instant invention. A non-peptide  
10 compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or  
15 pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of:  $-\text{CH}_2\text{NH}-$ ,  $-\text{CH}_2\text{S}-$ ,  $-\text{CH}_2-\text{CH}_2-$ ,  $-\text{CH}=\text{CH}-$  (cis and trans),  $-\text{COCH}_2-$ ,  $-\text{CH}(\text{OH})\text{CH}_2-$ , and  $-\text{CH}_2\text{SO}-$ , by methods well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of  
20 L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al.*, *Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which  
25 cyclize the peptide.

A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino  
30 acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid

sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, 5 more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

10 Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid

15 substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not

20 substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.). Proteins. Structures and Molecular Principles, W. H.

acids may also be suitable components for polypeptides of the present invention.

Examples of unconventional amino acids include: 4-hydroxyproline,  $\gamma$ -carboxyglutamate,

-N,N,N-trimethyllysine, -N-acetyllysine, O-phosphoserine, N-acetylserine,

N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other

5 similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a protein from another organism  
10 if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous,"  
15 this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence  
20 similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino  
25 acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino  
30 acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted

upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g., Pearson, Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are

5 conservative substitutions for one another:

- 1) Serine (S), Threonine (T);
- 2) Aspartic Acid (D), Glutamic Acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 10 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al., Science* 256: 1443-45 (1992), herein incorporated by reference. A "moderately conservative" replacement is  
15 any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid  
20 substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutin thereof. *See, e.g., GCG Version 6.1.* Other programs include FASTA, discussed *supra*.

25 A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g., Altschul et al., J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al., Nucleic Acids Res.* 25:3389-402 (1997); herein incorporated by reference. Preferred parameters for blastp are:

30       Expectation value:    10 (default)  
         Filter:               seg (default)  
         Cost to open a gap:  11 (default)



-28-

- Cost to extend a gap: 1 (default)  
Max. alignments: 100 (default)  
Word size: 11 (default)  
No. of descriptions: 100 (default)  
5 Penalty Matrix: BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number  
10 of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best  
15 overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

20 An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', F(ab')<sub>2</sub>, Fv,  
25 dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an F(ab')<sub>2</sub> fragment is a bivalent fragment comprising two Fab  
30 fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single

arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g., Ward et al., Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular  
5 species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made  
10 as a single protein chain. *See, e.g., Bird et al., Science* 242: 423-426 (1988); Huston *et al., Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain  
15 and creating two antigen binding sites. *See e.g., Holliger et al., Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al., Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may  
20 incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one  
25 binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-  
30 associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that

purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

5 A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist  
10 of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than 1  $\mu$ M, preferably less than 100 nM and most preferably less than 10 nM.

15 The term "patient" as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

20 The term "breast specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the breast as compared to other tissues in the body. In a preferred embodiment, a "breast specific" nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the "breast specific" nucleic acid molecule or polypeptide is  
25 expressed at a level that is 10-fold higher than any other tissue in the body, more preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels,  
30 such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

*Nucleic Acid Molecules*

5           One aspect of the invention provides isolated nucleic acid molecules that are specific to the breast or to breast cells or tissue or that are derived from such nucleic acid molecules. These isolated breast specific nucleic acids (BSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid  
10 molecule encodes a polypeptide that is specific to breast, a breast-specific polypeptide (BSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 160 through 282. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 159.

15           A BSNA may be derived from a human or from another animal. In a preferred embodiment, the BSNA is derived from a human or other mammal. In a more preferred embodiment, the BSNA is derived from a human or other primate. In an even more preferred embodiment, the BSNA is derived from a human.

          By "nucleic acid molecule" for purposes of the present invention, it is also meant  
20 to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding a BSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode a BSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a BSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively  
25 hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 160 through 282. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 159.

          In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a  
30 nucleic acid molecule encoding a BSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a BSP under moderate stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule

encoding a BSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 160 through 282. In a yet more preferred  
5 embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 159. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

10 By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding a BSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human BSP. In a more preferred embodiment, the nucleic acid molecule  
15 exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 160 through 282. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding a BSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 160 through 282, more preferably at least  
20 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a BSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid  
25 molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a BSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a BSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule  
30 comprising a nucleic acid sequence of SEQ ID NO: 1 through 159. In a preferred embodiment, the nucleic acid molecule is one that has at least 60% sequence identity with a BSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 159,

more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with a BSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a BSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one that exhibits sequence identity over its entire length to a BSNA or to a nucleic acid molecule encoding a BSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the BSNA or the nucleic acid molecule encoding a BSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 160 through 282 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 159.

The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the BSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of a BSNA. Further, the substantially

similar nucleic acid molecule may or may not be a BSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is a BSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of a BSNA or a nucleic acid encoding a BSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes a BSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a BSP comprising an amino acid sequence of SEQ ID NO: 160 through 282. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a BSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 159. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is a BSP. However, in a preferred embodiment, the part encodes a BSP. In one aspect, the invention comprises a part of a BSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a BSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of a BSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes a BSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides.

The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutein or a  
5 polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (e.g., reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular  
10 Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include  
15 nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-  
20 discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

25 In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

30 Common radiolabeled analogues include those labeled with  $^{33}\text{P}$ ,  $^{32}\text{P}$ , and  $^{35}\text{S}$ , such as  $^{-32}\text{P}$ -dATP,  $^{-32}\text{P}$ -dCTP,  $^{-32}\text{P}$ -dGTP,  $^{-32}\text{P}$ -dTTP,  $^{-32}\text{P}$ -3'dATP,  $^{-32}\text{P}$ -ATP,  $^{-32}\text{P}$ -CTP,  $^{-32}\text{P}$ -GTP,  $^{-32}\text{P}$ -UTP,  $^{-35}\text{S}$ -dATP,  $\alpha$ - $^{35}\text{S}$ -GTP,  $\alpha$ - $^{33}\text{P}$ -dATP, and the like.



Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. See Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label

5 (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see Alers et al., Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*, *BioTechniques* 16: 148-153. (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker

10 (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the

15 nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g., Tyagi et al., Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci.*

20 *USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517; 5,723,591 and 5,538,848; Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by

25 reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See Hartmann et al. (eds.), Manual of Antisense Methodology: Perspectives in*

*Antisense Science*, Kluwer Law International (1999); Stein *et al. (eds.), Applied*

30 *Antisense Oligonucleotide Technology*, Wiley-Liss (1998); Chadwick *et al. (eds.), Oligonucleotides as Therapeutic Agents - Symposium No. 209*, John Wiley & Son Ltd (1997); the disclosures of which are incorporated herein by reference in their entireties.

Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. See Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.

- Modified oligonucleotide backbones include, without limitation,
- 5 phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having
- 10 normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019;
- 15 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.
- 20 Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a
- 25 nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH<sub>2</sub> component parts. Representative U.S. patents
- 30 that teach the preparation of the above backbones include, but are not limited to, U.S. Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307;

5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (see, e.g., "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The  $T_m$  of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the  $T_m$  of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the  $T_m$  by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the  $T_m$  by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. See, e.g., Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1):

3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta*. 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure  
5 throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24:  
10 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed,  
15 partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by  
20 reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta*. 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

25

#### *Methods for Using Nucleic Acid Molecules as Probes and Primers*

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic  
30 acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of a BSNA, such as deletions, insertions, translocations, and duplications of the BSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.,*

5    Andreeff *et al.* (eds.), Introduction to Fluorescence *In Situ* Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.,* Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid

10   molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present

15   invention can be used as probes to detect, characterize, and quantify BSNA in, and isolate BSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A<sup>+</sup>-selected RNA samples. In another aspect, the isolated nucleic acid molecules of the

20   present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.,* Schwarchzacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization

25   probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to BSNA, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the

30   nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*;

Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used  
5 as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a BSP. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 160 through 282.  
10 In another preferred embodiment, the probe or primer is derived from a BSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 159.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17  
15 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using  
20 oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the  
25 art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques,  
30 John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of

which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books  
5 (1995); the disclosure of which is incorporated herein by reference in its entirety.

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion  
10 protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Patent 6,004,744,  
15 the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by  
20 reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-  
25 planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or  
30 positively-charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another



embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

#### *Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides*

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acids of the present invention, alone or as fusions to heterologous

polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are  
10 incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that  
15 expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or  
20 GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

25 In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial  
30 hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous

derivatives of phage lambda, *e.g.*, NM989,  $\lambda$ GT10 and  $\lambda$ GT11, and other phages, *e.g.*, M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline,  
5 chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous  
10 recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and  
15 a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEplac series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2 $\mu$  plasmids and derivatives thereof, and improved shuttle vectors such as those  
20 described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIp1ac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as *ura3-52*, *his3-D1*, *leu2-D1*, *trp1-D1* and *lys2-201*.

25 Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda*, *e.g.*, Sf9 and Sf21 cell lines, and expresSFT<sup>TM</sup> cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of  
30 interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between

these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally

contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the *trc* promoter, a hybrid derived from the *trp* and *lac* promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of *fd* coat protein, or the *araBAD* operon. Prokaryotic expression vectors may further include transcription terminators, such as the *aspA* terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the *CYC1* promoter, the *GAL1* promoter, the *GAL10* promoter, *ADH1* promoter, the promoters of the yeast  $\alpha$ -mating system, or the *GPD* promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the *CYC1* or *ADH1* gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the

Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from  
5 the gene comprising the BSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit  $\beta$ -globin gene and the SV40 splice elements.

10 Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell  
15 genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*,  
20 Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the *trc* promoter, which is regulated  
25 by the *lac* operon, and the *pL* promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid *Plac/ara-1* promoter and the *PLtetO-1* promoter. The *PLtetO-1* promoter takes advantage of the high expression levels  
30 from the *PL* promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the *Tn10* tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline

(Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone  
5 receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates  
10 purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™  
15 system, New England Biolabs, Inc., Beverly, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega,  
20 Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example,  
25 the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences  
30 that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that

carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the  $\alpha$ -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay<sup>TM</sup> vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from



those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. See Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g.* Cormack *et al.*, *Gene* 173: 33-38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g.* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.* Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entirety. See also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable

expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The *bsd* gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

10 Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™ 293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, 15 Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts 20 function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the 25 ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described 30 above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation,

and acylation, and it is an aspect of the present invention to provide BSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation, acetylation, prenylation, methylation, acetylation, arginylation, ubiquination and racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., [www.expasy.org](http://www.expasy.org) (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in web sites such as the Delta Mass database <http://www.abrf.org/ABRF/Research/Committees/deltamass/deltamass.html> (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. *Nucleic Acids Res.* 29; 332-335 (2001) and <http://www.glycosuite.com/> (accessed October 19, 2001); "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. *Nucleic Acids Research*, 27: 370-372 (1999) and <http://www.cbs.dtu.dk/databases/OGLYCBASE/> (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. *Nucleic Acids Res* 27(1):237-239 (1999) and <http://www.cbs.dtu.dk/databases/PhosphoBase/> (accessed October 19, 2001); or <http://pir.georgetown.edu/pirwww/search/textresid.html> (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur

in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g., p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the

desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one  
5 may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website [www.expasy.org](http://www.expasy.org). The nucleic acid molecule is then be introduced into a host cell that is  
10 capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

15 In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the  
20 product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression  
25 vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have  
30 biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors

typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell  
5 (e.g., conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an  
10 expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be  
15 able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera*  
20 *frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from  
25 *Spodoptera frugiperda*, e.g., Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3  
30 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and

readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from breast are particularly preferred because they may provide a more native  
5 post-translational processing. Particularly preferred are human breast cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number  
10 of texts and laboratory manuals in the art. *See, e.g.,* Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend  
15 primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (*e.g.,* Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

20 Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, *e.g.,* with  $\text{CaCl}_2$ , or a solution of  $\text{Mg}^{2+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Rb}^+$  or  $\text{K}^+$ , dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent  
25 strains are also available commercially (*e.g.,* Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5 competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent *E. coli* Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse  
30 treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols



(BioRad, Richmond, CA, USA) ([http://www.biorad.com/LifeScience/pdf/New\\_Gene\\_Pulser.pdf](http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf)).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the  
5 action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and  $\text{Ca}^{2+}$ . Subsequently, the cells are resuspended in a solution of sorbitol, mixed  
10 with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective  
15 medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension  
20 pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

25 Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with  $\text{CaPO}_4$  or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for  $\text{CaPO}_4$  transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated  
30 transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent,

FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA).

Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (<http://www.bio-rad.com/LifeScience/pdf/>

- 5 New\_Gene\_Pulser.pdf); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.,* Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA*  
10 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

- Purification of recombinantly expressed proteins is now well by those skilled in the art. *See, e.g.,* Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid  
15 Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.),  
20 Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

- Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by  
25 means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

### Polypeptides

- 30 Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is a breast specific polypeptide (BSP). In an even more preferred embodiment, the

polypeptide is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 160 through 282. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification  
5 and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide, wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of a BSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 160  
10 through 282. A polypeptide that comprises only a fragment of an entire BSP may or may not be a polypeptide that is also a BSP. For instance, a full-length polypeptide may be breast-specific, while a fragment thereof may be found in other tissues as well as in breast. A polypeptide that is not a BSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing  
15 animals to prepare anti-BSP antibodies. However, in a preferred embodiment, the part or fragment is a BSP. Methods of determining whether a polypeptide is a BSP are described *infra*.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA*  
20 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

25 Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. *See, e.g., Lerner, Nature* 299: 592-596 (1982); Shinnick *et al., Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al., Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further  
30 described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for

the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies  
5 (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

10 The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75  
15 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, *e.g.*, a BSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically  
20 synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See, e.g.*, Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a  
25 fragment of polypeptide of the invention, preferably a BSP, may be produced by chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably a BSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants,  
30 fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be breast-specific. In a preferred embodiment, the mutein is breast-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 160 through 282. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 160 through 282. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 160 through 282.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is breast-specific, as described below. Multiple random mutations can be introduced into the gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo*

mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.,* Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel  
5 (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the polypeptide is homologous to a BSP. In an even more preferred embodiment, the  
10 polypeptide is homologous to a BSP selected from the group having an amino acid sequence of SEQ ID NO: 160 through 282. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to a BSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity to an comprising an amino acid sequence of SEQ ID NO: 160 through 282. In an even  
15 more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 160 through 282. In a yet more preferred embodiment, the homologous polypeptide is one that exhibits at least 85%,  
20 more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 160 through 282. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to a BSP comprising an amino acid sequence  
25 of SEQ ID NO: 160 through 282. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to a BSNA. In a preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a  
30 BSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the BSNA is selected from the group consisting of SEQ ID NO: 1 through 159. In another preferred embodiment, the

homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a BSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the BSP is selected from the group consisting of SEQ ID NO: 160 through 282.

- 5           The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 160 through 282. The homologous polypeptide may also be a naturally-  
10   occurring polypeptide from a human, when the BSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring  
15   polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. In another embodiment, the homologous polypeptide may be  
20   one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a BSP. Further, the homologous protein may or may not encode polypeptide that is a BSP. However, in a  
25   preferred embodiment, the homologous polypeptide encodes a polypeptide that is a BSP.

- Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but  
30   also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the

present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding a BSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 160 through 282. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 159.

In another embodiment, the invention provides polypeptides which comprise derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a BSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 160 through 282, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as  $^{125}\text{I}$ ,  $^{32}\text{P}$ ,  $^{35}\text{S}$ , and  $^3\text{H}$ . In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983);



Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591,

BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

- 5           The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS,
- 10 HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC,
- 15 SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

- The polypeptides, fragments, and fusion proteins of the present invention can be
- 20 conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

- The polypeptides, fragments, and fusion proteins of the present invention can also
- 25 usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-BSP antibodies.

- The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum
- 30 half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999),

incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

- 5           In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a BSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 160 through 282. In a preferred embodiment, the analog is one that comprises one or
- 10 more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to a BSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH<sub>2</sub>NH--, --CH<sub>2</sub>S--, --CH<sub>2</sub>-CH<sub>2</sub>--,
- 15 --CH=CH--(cis and trans), --COCH<sub>2</sub>--, --CH(OH)CH<sub>2</sub>-- and --CH<sub>2</sub>SO--. In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of a BSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis; peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific
- 20 three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g., Kole et al., Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.
- 25           Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000);
- 30 Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer

Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entirety.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added  
5 using biotinoyl-(9-fluorenylmethoxycarbonyl)-L-lysine (FMOC biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The FMOC and *t*BOC derivatives of dabcyL-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyL chromophore at selected sites in the peptide sequence during  
10 synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyL quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS-FMOC-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be  
15 incorporated during automated FMOC synthesis of peptides using (FMOC)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side  
20 chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other FMOC-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-  
25 aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic  
30 acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-

- 2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)- $\beta$ -alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphthoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

- Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

### *Fusion Proteins*

- The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is a BSP. In a more preferred embodiment, the polypeptide that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 160 through 282, or is a mutein, homologous polypeptide,

analog or derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NO: 1 through 159, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 159.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g., Ausubel, Chapter 16, (1992), supra.* Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation

- of secretion signals and/or leader sequences. For example, a His<sup>6</sup> tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.
- 10        Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. See Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabrizio *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-
- 20        hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. *Proc Natl Acad Sci U S A* 95, 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein
- 25        interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by
- 30

reference in their entirety. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded  
5 protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides and fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin  
10 A, *anthrax* toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins,  $\beta$ -galactosidase, biotin trpE, protein A,  $\beta$ -lactamase, -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine  
15 at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast \_ mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. See, e.g., Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by  
20 enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (e.g., a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind  
25 the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the BSP.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize BSPs, their allelic  
30 variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly BSPs, e.g. by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser



scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of BSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of BSPs.

- 5           One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon  
10 linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-  
15 102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

- Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of one having ordinary skill in the art. *See, e.g.*, Scopes, Protein Purification, 2d ed. (1987).  
20 Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

- Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-  
25 proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

- Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For  
30 example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

### Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that is a BSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 160 through 282, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a BSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or visa versa. In addition, alternative splice forms of a BSP may be indicative of cancer. Differential degradation of the C or N-terminus of a BSP may also be a marker or target for anticancer therapy. For example, a BSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-BSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the protein of the present invention in samples derived from human breast.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the

present invention will be at least about  $1 \times 10^{-6}$  molar (M), typically at least about  $5 \times 10^{-7}$  M,  $1 \times 10^{-7}$  M, with affinities and avidities of at least  $1 \times 10^{-8}$  M,  $5 \times 10^{-9}$  M,  $1 \times 10^{-10}$  M and up to  $1 \times 10^{-13}$  M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as  
5 IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such  
10 antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic  
15 mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by  
20 reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered  
25 antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger  
30 mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are

hereby incorporated in their entirety. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

5 As discussed above, virtually all fragments of 8 or more contiguous amino acids of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

10 Immunogenicity can also be conferred by fusion of the polypeptide and fragments of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725  
15 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. See Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by reference. Immunization protocols often include multiple immunizations, either with or  
20 without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990)).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the proteins of the present invention and monoclonal  
30 antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular

advantage in detection of the proteins of the present invention, in human serum or tissues (Vikinge et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998)).

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, 5 *supra*; Zola, *supra*; Howard et al. (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their 10 entireties, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding 15 antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S. Patent 20 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody 25 fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein 30 (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. See, *e.g.*, Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths et al., *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom et al., *Immunotechnology*,

4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujaime *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994).

- 5 Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.*, Barbas (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.

Typically, phage-displayed antibody fragments are scFv fragments or Fab  
10 fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

- 15 For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.*, Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, *J. Biotechnol.* 76(2-3): 157-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein  
20 by reference in their entireties.

- Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.*, Li *et al.*, *Protein Expr. Purif.*  
25 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

- Antibodies and fragments and derivatives thereof of the present invention can  
30 also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*,

*Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entirety.

5           Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock *et al.*, *J. Immunol. Methods.* 231: 147-57 (1999); Young *et al.*, *Res. Immunol.* 149: 609-10 (1998); Limonta *et al.*, *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entirety.

10           Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

            Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression  
15           systems for expression of antibodies.

            Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem. (Tokyo)* 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57  
20           (1999), the disclosures of which are incorporated herein by reference in their entirety.

            The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the  
25           proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

            Among such useful fragments are Fab, Fab', Fv, F(ab')<sub>2</sub>, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

30           It is also an aspect of the present invention to provide antibody derivatives that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated



nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

5           Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

10           Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA* 81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature*  
15 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2  
20 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain  
25 diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof,  
30 whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions

including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase,  $\beta$ -galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopyranoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide ( $H_2O_2$ ), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate

reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. *See, e.g., Thorpe et al., Methods Enzymol.* 133: 331-53 (1986); Kricka *et al., J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al., J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

10 As another example, when the antibodies of the present invention are used, *e.g.,* for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

15 For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

20 Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568,  
25 BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5,  
30 Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, e.g., for Western blotting applications, they can usefully be labeled with radioisotopes, such as  $^{33}\text{P}$ ,  $^{32}\text{P}$ ,  $^{35}\text{S}$ ,  $^3\text{H}$ ,  
5 and  $^{125}\text{I}$ .

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be  $^{228}\text{Th}$ ,  $^{227}\text{Ac}$ ,  $^{225}\text{Ac}$ ,  $^{223}\text{Ra}$ ,  $^{213}\text{Bi}$ ,  $^{212}\text{Pb}$ ,  $^{212}\text{Bi}$ ,  $^{211}\text{At}$ ,  $^{203}\text{Pb}$ ,  $^{194}\text{Os}$ ,  $^{188}\text{Re}$ ,  $^{186}\text{Re}$ ,  $^{153}\text{Sm}$ ,  $^{149}\text{Tb}$ ,  $^{131}\text{I}$ ,  $^{125}\text{I}$ ,  $^{111}\text{In}$ ,  $^{105}\text{Rh}$ ,  $^{99\text{m}}\text{Tc}$ ,  $^{97}\text{Ru}$ ,  $^{90}\text{Y}$ ,  $^{90}\text{Sr}$ ,  $^{88}\text{Y}$ ,  $^{72}\text{Se}$ ,  $^{67}\text{Cu}$ , or  $^{47}\text{Sc}$ .

10 As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the  
15 application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, *diphtheria*  
20 toxin, *shiga* toxin A, *anthrax* toxin lethal factor, or ricin. See Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate,  
25 and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more  
30 of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to  
5 paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in  
10 prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind  
15 specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present  
20 invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody  
25 molecule, or to alter it in any other way that may render it more suitable for a particular application.

#### Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human  
30 organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding a BSP. In a preferred embodiment, the BSP comprises an amino acid

sequence selected from SEQ ID NO: 160 through 282, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a BSNA of the invention, preferably a BSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID  
5 NO: 1 through 159, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human BSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-  
10 human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g.,* Hogan *et al.*, Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology:  
15 A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.,* Paterson *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11:  
20 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989 retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g.,* Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g.,* Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g.,* Lo, 1983, *Mol. Cell. Biol.*  
25 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g.,* Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g.,* Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of  
30 nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.,* Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e., a*

nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene  
5 may also be selectively introduced into and activated in a particular cell type by following, e.g., the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

10 Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using  
15 techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

20 Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels  
25 because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is  
30 appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of

the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g., Gu et al., Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g., Smithies et al., Nature* 317: 230-234 (1985); Thomas *et al., Cell* 51: 503-512 (1987); Thompson *et al., Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g., Thomas, supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g., knockouts*) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC



compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt  
5 the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

10 The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

15 Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

20 When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the  
25 introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such  
30 conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 159 and SEQ ID NO: 160 through 282 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set

representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

#### Diagnostic Methods for Breast Cancer

30

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by

comparing expression of a BSNA or a BSP in a human patient that has or may have breast cancer, or who is at risk of developing breast cancer, with the expression of a BSNA or a BSP in a normal human control. For purposes of the present invention, “expression of a BSNA” or “BSNA expression” means the quantity of BSG mRNA that  
5 can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term “expression of a BSP” or “BSP expression” means the amount of BSP that can be measured by any method known in the art or the level of translation of a BSG BSNA that can be measured by any method known in the art.

10 The present invention provides methods for diagnosing breast cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of BSNA or BSP in cells, tissues, organs or bodily fluids compared with levels of BSNA or BSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of a BSNA or BSP in  
15 the patient versus the normal human control is associated with the presence of breast cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing breast cancer in a patient by analyzing changes in the structure of the mRNA of a BSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations in  
20 polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing breast cancer in a patient by analyzing changes in a BSP compared to a BSP from a normal control. These changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the BSP or subcellular BSP localization.

25 In a preferred embodiment, the expression of a BSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from SEQ ID NO: 160 through 282, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the BSNA expression that is measured is the level of expression of a BSNA mRNA selected from SEQ ID NO: 1 through 159, or a  
30 hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. BSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by

Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.,* Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. BSNA transcription may be measured by any method known in the art including using a reporter  
5 gene hooked up to the promoter of a BSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.,* aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, BSNA expression may be compared to a known control, such as normal breast nucleic acid, to detect a change in expression.

10 In another preferred embodiment, the expression of a BSP is measured by determining the level of a BSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 160 through 282, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for  
15 instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of BSNA or BSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of breast cancer. The expression level of a BSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the BSP expression level may be  
20 determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See, e.g.,* Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the BSP  
25 structure may be determined by any method known in the art, including, *e.g.,* using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An  
30 antibody specific to a BSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-BSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a

protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the BSP will bind to the anti-BSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-BSP antibody that is linked to a detectable reagent (a radioactive  
5 substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the BSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of a BSP in the sample. For an RIA, the solid support is counted for radioactive  
10 decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure BSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-BSP antibody is attached to a solid support and an allocated amount of a labeled BSP and a sample of interest are incubated  
15 with the solid support. The amount of labeled BSP detected which is attached to the solid support can be correlated to the quantity of a BSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides  
20 are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with  
25 chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of a BSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant  
30 cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other

mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more BSNAs of interest. In this approach, all or a portion of one or more BSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of BSNA or BSP includes, without limitation, breast tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, breast cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary breast cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and colon. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediastinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. See Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in BSNAs or BSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of a BSNA or BSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other BSNA or BSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular BSNA or BSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

#### *Diagnosing*

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more BSNAs and/or BSPs in a sample from a patient suspected of having breast cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a BSNA and/or BSP and then ascertaining whether the patient has breast cancer from the expression level of the BSNA or BSP. In general, if high expression relative to a control of a BSNA or BSP is indicative of breast cancer, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of breast cancer, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether breast cancer has metastasized in a patient. One may identify whether the breast cancer has metastasized by measuring the expression levels and/or structural alterations of one or more BSNAs and/or BSPs in a variety of tissues. The presence of a BSNA or BSP in a



certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of a BSNA or BSP is associated with breast cancer. Similarly, the presence of a BSNA or BSP in a tissue at levels lower than that of corresponding noncancerous tissue is  
5 indicative of metastasis if low level expression of a BSNA or BSP is associated with breast cancer. Further, the presence of a structurally altered BSNA or BSP that is associated with breast cancer is also indicative of metastasis.

In general, if high expression relative to a control of a BSNA or BSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of  
10 the BSNA or BSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the BSNA or BSP is at least two times  
15 lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The BSNA or BSP of this invention may be used as element in an array or a multi-analyte test to recognize expression patterns associated with breast cancers or other  
20 breast related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of breast disorders.

### *Staging*

25 The invention also provides a method of staging breast cancer in a human patient. The method comprises identifying a human patient having breast cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more BSNAs or BSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the  
30 expression level of one or more BSNAs or BSPs is determined for each stage to obtain a standard expression level for each BSNA and BSP. Then, the BSNA or BSP expression levels are determined in a biological sample from a patient whose stage of cancer is not

known. The BSNA or BSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the BSNA and BSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a BSNA or  
5 BSP to determine the stage of a breast cancer.

### *Monitoring*

Further provided is a method of monitoring breast cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient  
10 to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the breast cancer. The method comprises identifying a human patient that one wants to monitor for breast cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for  
15 expression levels of one or more BSNA or BSPs, and comparing the BSNA or BSP levels over time to those BSNA or BSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a BSNA or BSP that are associated with breast cancer.

If increased expression of a BSNA or BSP is associated with metastasis,  
20 treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of a BSNA or BSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or  
25 failure to progress to a neoplastic lesion. If decreased expression of a BSNA or BSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of a BSNA or BSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of BSNA or BSPs are  
30 determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of breast cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a BSNA and/or BSP. The present invention provides a method in which a test sample is obtained from a human patient and one or  
5 more BSNAs and/or BSPs are detected. The presence of higher (or lower) BSNA or BSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly breast cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more BSNAs and/or BSPs of the invention can also be monitored by analyzing levels of expression of the  
10 BSNAs and/or BSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

#### *Detection of Genetic Lesions or Mutations*

15 The methods of the present invention can also be used to detect genetic lesions or mutations in a BSG, thereby determining if a human with the genetic lesion is susceptible to developing breast cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing breast cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of  
20 one or more nucleotides from the BSGs of this invention, a chromosomal rearrangement of BSG, an aberrant modification of BSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a BSG. Methods to detect such lesions in the BSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

#### 25 Methods of Detecting Noncancerous Breast Diseases

The invention also provides a method for determining the expression levels and/or structural alterations of one or more BSNAs and/or BSPs in a sample from a patient suspected of having or known to have a noncancerous breast disease. In general,  
30 the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of a BSNA and/or BSP, comparing the expression level or structural alteration of the BSNA or BSP to a normal breast control,

and then ascertaining whether the patient has a noncancerous breast disease. In general, if high expression relative to a control of a BSNA or BSP is indicative of a particular noncancerous breast disease, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times higher, and more preferably are at  
5 least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of a noncancerous breast disease, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times lower, more preferably are at least five times lower, even  
10 more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether a BSNA and/or BSP is associated with a particular noncancerous breast disease by obtaining breast tissue  
15 from a patient having a noncancerous breast disease of interest and determining which BSNAs and/or BSPs are expressed in the tissue at either a higher or a lower level than in normal breast tissue. In another embodiment, one may determine whether a BSNA or BSP exhibits structural alterations in a particular noncancerous breast disease state by obtaining breast tissue from a patient having a noncancerous breast disease of interest  
20 and determining the structural alterations in one or more BSNAs and/or BSPs relative to normal breast tissue.

#### Methods for Identifying Breast Tissue

25 In another aspect, the invention provides methods for identifying breast tissue. These methods are particularly useful in, *e.g.*, forensic science, breast cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is breast tissue or has breast tissue-like characteristics. The method comprises the  
30 steps of providing a sample suspected of comprising breast tissue or having breast tissue-like characteristics, determining whether the sample expresses one or more BSNAs and/or BSPs, and, if the sample expresses one or more BSNAs and/or BSPs, concluding that the sample comprises breast tissue. In a preferred embodiment, the BSNA encodes a

polypeptide having an amino acid sequence selected from SEQ ID NO: 160 through 282, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the BSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 159, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses a BSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether a BSP is expressed. Determining whether a sample expresses a BSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the BSP has an amino acid sequence selected from SEQ ID NO: 160 through 282, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two BSNAs and/or BSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five BSNAs and/or BSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is breast tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into breast tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new breast tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

#### Methods for Producing and Modifying Breast Tissue

In another aspect, the invention provides methods for producing engineered breast tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing a BSNA or a BSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of breast tissue cells. In a preferred

embodiment, the cells are pluripotent. As is well-known in the art, normal breast tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered breast tissue or cells comprises one of these cell types. In another embodiment, the engineered breast tissue or cells comprises more than one breast cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the breast cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more BSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode BSPs having amino acid sequences selected from SEQ ID NO: 160 through 282, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 159, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, a BSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial breast tissue may be used to treat patients who have lost some or all of their breast function.

#### Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises a BSNA or part thereof. In a more preferred embodiment, the BSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 159, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises a BSP or fragment thereof. In a more preferred embodiment, the BSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 160 through 282, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the

pharmaceutical composition comprises an anti-BSP antibody, preferably an antibody that specifically binds to a BSP having an amino acid that is selected from the group consisting of SEQ ID NO: 160 through 282, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the  
5 polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in  
10 Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20<sup>th</sup> ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7<sup>th</sup> ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3<sup>rd</sup> ed. (2000), the disclosures of which are incorporated herein by reference in their entireties,  
15 and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation,  
20 topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain  
25 suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium  
30 carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

- 5           Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

- 10           Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

- Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.
- 15

- Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.
- 20

- Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.
- 25

- Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.
- 30



The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

10 Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil  
15 base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of  
25 the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature  
30 of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot

injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically.

5 For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration  
10 of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

15 For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as  
20 cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for  
25 treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the  
30 present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts

tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

5       The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example BSP polypeptide, fusion protein, or fragments thereof, antibodies specific for  
10   BSP, agonists, antagonists or inhibitors of BSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed  
15   by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in  
20   one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of  
25   circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

30       The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age,

-111-

weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

- 5           Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be  
10 administered in multiple doses per day, if desired, to achieve the total desired daily dose.

- Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells,  
15 conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

## 20   Therapeutic Methods

- The present invention further provides methods of treating subjects having defects in a gene of the invention, *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of breast function. As used herein,  
25 "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

### *Gene Therapy and Vaccines*

- 30           The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication

incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further  
5 described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g., Doronin et al., J. Virol.* 75: 3314-24 (2001).

10 In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of a BSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a BSP are  
15 administered, for example, to complement a deficiency in the native BSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g., Cid-Arregui, supra.* In a preferred embodiment, the nucleic acid molecule encodes a BSP having the amino acid sequence of SEQ ID NO: 160 through  
20 282, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a BSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in BSP  
25 production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a BSP having the amino acid sequence of SEQ ID NO: 160 through 282, or a fragment, fusion protein, allelic variant or homolog thereof.

#### *Antisense Administration*

Antisense nucleic acid compositions, or vectors that drive expression of a BSP  
30 antisense nucleic acid, are administered to downregulate transcription and/or translation of a BSP in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a BSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

5 Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to BSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are  
10 incorporated herein by reference in their entireties.

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the BSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9  
15 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid  
20 molecule encoding a BSP, preferably a BSP comprising an amino acid sequence of SEQ ID NO: 160 through 282, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 159, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

## 25 *Polypeptide Administration*

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a BSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant BSP defect.

30 Protein compositions are administered, for example, to complement a deficiency in native BSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to BSP. The immune response can

be used to modulate activity of BSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate BSP.

- 5           In a preferred embodiment, the polypeptide is a BSP comprising an amino acid sequence of SEQ ID NO: 160 through 282, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 159, or a part, allelic variant, substantially similar or hybridizing nucleic acid  
10 thereof.

*Antibody, Agonist and Antagonist Administration*

- In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is  
15 administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of BSP, or to target therapeutic agents to sites of BSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a BSP comprising an amino acid sequence of SEQ ID NO: 160 through 282, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred  
20 embodiment, the antibody specifically binds to a BSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 159, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

          The present invention also provides methods for identifying modulators which bind to a BSP or have a modulatory effect on the expression or activity of a BSP.

- 25 Modulators which decrease the expression or activity of BSP (antagonists) are believed to be useful in treating breast cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of a BSP can also be designed, synthesized and tested for use in the imaging and treatment of breast  
30 cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the BSPs identified herein. Molecules identified in the library as being capable of binding to a BSP are key candidates for

further evaluation for use in the treatment of breast cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a BSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of BSP is  
5 administered. Antagonists of BSP can be produced using methods generally known in the art. In particular, purified BSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a BSP.

In other embodiments a pharmaceutical composition comprising an agonist of a  
10 BSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a BSP comprising an amino acid sequence of SEQ  
ID NO: 160 through 282, or a fusion protein, allelic variant, homolog, analog or  
15 derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a BSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 159, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

#### *Targeting Breast Tissue*

20 The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the breast or to specific cells in the breast. In a preferred embodiment, an anti-BSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if breast tissue needs to be  
25 selectively destroyed. This would be useful for targeting and killing breast cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting breast cell function.

In another embodiment, an anti-BSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be  
30 useful for determining and monitoring breast function, identifying breast cancer tumors, and identifying noncancerous breast diseases.



## EXAMPLES

### Example 1: Gene Expression analysis

BSGs were identified by a systematic analysis of gene expression data in the LIFESEQ® Gold database available from Incyte Genomics Inc (Palo Alto, CA) using the data mining software package CLASP™ (Candidate Lead Automatic Search Program). CLASP™ is a set of algorithms that interrogate Incyte's database to identify genes that are both specific to particular tissue types as well as differentially expressed in tissues from patients with cancer. LifeSeq® Gold contains information about which genes are expressed in various tissues in the body and about the dynamics of expression in both normal and diseased states. CLASP™ first sorts the LifeSeq® Gold database into defined tissue types, such as breast, ovary and prostate. CLASP™ categorizes each tissue sample by disease state. Disease states include "healthy," "cancer," "associated with cancer," "other disease" and "other." Categorizing the disease states improves our ability to identify tissue and cancer-specific molecular targets. CLASP™ then performs a simultaneous parallel search for genes that are expressed both (1) selectively in the defined tissue type compared to other tissue types and (2) differentially in the "cancer" disease state compared to the other disease states affecting the same, or different, tissues. This sorting is accomplished by using mathematical and statistical filters that specify the minimum change in expression levels and the minimum frequency that the differential expression pattern must be observed across the tissue samples for the gene to be considered statistically significant. The CLASP™ algorithm quantifies the relative abundance of a particular gene in each tissue type and in each disease state.

To find the BSGs of this invention, the following specific CLASP™ profiles were utilized: tissue-specific expression (CLASP 1), detectable expression only in cancer tissue (CLASP 2), highest differential expression for a given cancer (CLASP 4); differential expression in cancer tissue (CLASP 5), and. cDNA libraries were divided into 60 unique tissue types (early versions of LifeSeq® had 48 tissue types). Genes or ESTs were grouped into "gene bins," where each bin is a cluster of sequences grouped together where they share a common contig. The expression level for each gene bin was calculated for each tissue type. Differential expression significance was calculated with rigorous statistical significant testing taking into account variations in sample size and

relative gene abundance in different libraries and within each library (for the equations used to determine statistically significant expression see Audic and Claverie "The significance of digital gene expression profiles," Genome Res 7(10): 986-995 (1997), including Equation 1 on page 987 and Equation 2 on page 988, the contents of which are  
 5 incorporated by reference). Differentially expressed tissue-specific genes were selected based on the percentage abundance level in the targeted tissue versus all the other tissues (tissue-specificity). The expression levels for each gene in libraries of normal tissues or non-tumor tissues from cancer patients were compared with the expression levels in tissue libraries associated with tumor or disease (cancer-specificity). The results were  
 10 analyzed for statistical significance.

The selection of the target genes meeting the rigorous CLASP™ profile criteria were as follows:

- (a) CLASP 1: tissue-specific expression: To qualify as a CLASP 1 candidate, a gene must exhibit statistically significant expression in the tissue of interest compared to all other tissues. Only if the gene exhibits such differential  
 15 expression with a 90% of confidence level is it selected as a CLASP 1 candidate.
- (b) CLASP 2: detectable expression only in cancer tissue: To qualify as a CLASP 2 candidate, a gene must exhibit detectable expression in tumor tissues and undetectable expression in libraries from normal individuals and libraries from normal tissue obtained from diseased patients. In addition, such a gene must also exhibit further specificity for the tumor tissues of interest.  
 20
- (c) CLASP 5: differential expression in cancer tissue: To qualify as a CLASP 5 candidate, a gene must be differentially expressed in tumor libraries in the tissue of interest compared to normal libraries for all tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it  
 25 selected as a CLASP 5 candidate.

The CLASP™ scores for SEQ ID NO: 1-159 are listed below:

30	SEQ ID NO: 1	DEX0249_1	CLASP2
	SEQ ID NO: 2	DEX0249_2	CLASP2
	SEQ ID NO: 3	DEX0249_3	CLASP2
	SEQ ID NO: 4	DEX0249_4	CLASP2

-118-

	SEQ ID NO: 5	DEX0249_5	CLASP2
	SEQ ID NO: 6	DEX0249_6	CLASP2
	SEQ ID NO: 7	DEX0249_7	CLASP2
	SEQ ID NO: 8	DEX0249_8	CLASP2
5	SEQ ID NO: 9	DEX0249_9	CLASP2
	SEQ ID NO: 11	DEX0249_11	CLASP2
	SEQ ID NO: 12	DEX0249_12	CLASP2
	SEQ ID NO: 13	DEX0249_13	CLASP2
	SEQ ID NO: 14	DEX0249_14	CLASP2
10	SEQ ID NO: 15	DEX0249_15	CLASP2 CLASP1
	SEQ ID NO: 16	DEX0249_16	CLASP2 CLASP1
	SEQ ID NO: 17	DEX0249_17	CLASP2
	SEQ ID NO: 18	DEX0249_18	CLASP2
	SEQ ID NO: 19	DEX0249_19	CLASP2
15	SEQ ID NO: 20	DEX0249_20	CLASP2
	SEQ ID NO: 21	DEX0249_21	CLASP2
	SEQ ID NO: 22	DEX0249_22	CLASP2
	SEQ ID NO: 23	DEX0249_23	CLASP2
	SEQ ID NO: 24	DEX0249_24	CLASP2
20	SEQ ID NO: 25	DEX0249_25	CLASP2 CLASP1
	SEQ ID NO: 26	DEX0249_26	CLASP2 CLASP1
	SEQ ID NO: 27	DEX0249_27	CLASP2
	SEQ ID NO: 28	DEX0249_28	CLASP5 CLASP1
	SEQ ID NO: 29	DEX0249_29	CLASP2
25	SEQ ID NO: 30	DEX0249_30	CLASP2
	SEQ ID NO: 31	DEX0249_31	CLASP2
	SEQ ID NO: 32	DEX0249_32	CLASP2
	SEQ ID NO: 33	DEX0249_33	CLASP2
	SEQ ID NO: 34	DEX0249_34	CLASP2
30	SEQ ID NO: 35	DEX0249_35	CLASP2
	SEQ ID NO: 36	DEX0249_36	CLASP2
	SEQ ID NO: 37	DEX0249_37	CLASP2
	SEQ ID NO: 38	DEX0249_38	CLASP2
	SEQ ID NO: 39	DEX0249_39	CLASP2
35	SEQ ID NO: 40	DEX0249_40	CLASP2
	SEQ ID NO: 41	DEX0249_41	CLASP2
	SEQ ID NO: 42	DEX0249_42	CLASP2
	SEQ ID NO: 43	DEX0249_43	CLASP2
	SEQ ID NO: 44	DEX0249_44	CLASP2
40	SEQ ID NO: 45	DEX0249_45	CLASP2
	SEQ ID NO: 46	DEX0249_46	CLASP2
	SEQ ID NO: 47	DEX0249_47	CLASP2
	SEQ ID NO: 48	DEX0249_48	CLASP2
	SEQ ID NO: 49	DEX0249_49	CLASP2
45	SEQ ID NO: 50	DEX0249_50	CLASP2
	SEQ ID NO: 51	DEX0249_51	CLASP2
	SEQ ID NO: 52	DEX0249_52	CLASP2
	SEQ ID NO: 53	DEX0249_53	CLASP2

-119-

	SEQ ID NO: 54	DEX0249_54	CLASP2
	SEQ ID NO: 55	DEX0249_55	CLASP2
	SEQ ID NO: 56	DEX0249_56	CLASP2
	SEQ ID NO: 57	DEX0249_57	CLASP2
5	SEQ ID NO: 58	DEX0249_58	CLASP2
	SEQ ID NO: 59	DEX0249_59	CLASP2
	SEQ ID NO: 60	DEX0249_60	CLASP2
	SEQ ID NO: 61	DEX0249_61	CLASP2
	SEQ ID NO: 62	DEX0249_62	CLASP5 CLASP1
10	SEQ ID NO: 63	DEX0249_63	CLASP5 CLASP1
	SEQ ID NO: 64	DEX0249_64	CLASP5 CLASP1
	SEQ ID NO: 65	DEX0249_65	CLASP5 CLASP1
	SEQ ID NO: 66	DEX0249_66	CLASP2
	SEQ ID NO: 67	DEX0249_67	CLASP2
15	SEQ ID NO: 68	DEX0249_68	CLASP5 CLASP1
	SEQ ID NO: 69	DEX0249_69	CLASP2
	SEQ ID NO: 70	DEX0249_70	CLASP2
	SEQ ID NO: 71	DEX0249_71	CLASP2
	SEQ ID NO: 72	DEX0249_72	CLASP2
20	SEQ ID NO: 73	DEX0249_73	CLASP2
	SEQ ID NO: 74	DEX0249_74	CLASP2
	SEQ ID NO: 75	DEX0249_75	CLASP2
	SEQ ID NO: 76	DEX0249_76	CLASP2
	SEQ ID NO: 77	DEX0249_77	CLASP2
25	SEQ ID NO: 78	DEX0249_78	CLASP5 CLASP1
	SEQ ID NO: 79	DEX0249_79	CLASP2
	SEQ ID NO: 80	DEX0249_80	CLASP2
	SEQ ID NO: 81	DEX0249_81	CLASP2 CLASP1
	SEQ ID NO: 82	DEX0249_82	CLASP2
30	SEQ ID NO: 83	DEX0249_83	CLASP2
	SEQ ID NO: 84	DEX0249_84	CLASP2
	SEQ ID NO: 85	DEX0249_85	CLASP2
	SEQ ID NO: 86	DEX0249_86	CLASP2
	SEQ ID NO: 87	DEX0249_87	CLASP2
35	SEQ ID NO: 88	DEX0249_88	CLASP2
	SEQ ID NO: 89	DEX0249_89	CLASP2
	SEQ ID NO: 90	DEX0249_90	CLASP2 CLASP1
	SEQ ID NO: 91	DEX0249_91	CLASP2
	SEQ ID NO: 92	DEX0249_92	CLASP5 CLASP1
40	SEQ ID NO: 93	DEX0249_93	CLASP2
	SEQ ID NO: 94	DEX0249_94	CLASP2
	SEQ ID NO: 95	DEX0249_95	CLASP2
	SEQ ID NO: 96	DEX0249_96	CLASP2
	SEQ ID NO: 97	DEX0249_97	CLASP2
45	SEQ ID NO: 98	DEX0249_98	CLASP2
	SEQ ID NO: 99	DEX0249_99	CLASP2
	SEQ ID NO: 100	DEX0249_100	CLASP2
	SEQ ID NO: 101	DEX0249_101	CLASP2

-120-

	SEQ ID NO: 102	DEX0249_102	CLASP2
	SEQ ID NO: 103	DEX0249_103	CLASP2
	SEQ ID NO: 104	DEX0249_104	CLASP2
	SEQ ID NO: 105	DEX0249_105	CLASP5 CLASP1
5	SEQ ID NO: 106	DEX0249_106	CLASP5 CLASP1
	SEQ ID NO: 107	DEX0249_107	CLASP5 CLASP1
	SEQ ID NO: 108	DEX0249_108	CLASP5 CLASP1
	SEQ ID NO: 109	DEX0249_109	CLASP5 CLASP1
	SEQ ID NO: 110	DEX0249_110	CLASP2
10	SEQ ID NO: 111	DEX0249_111	CLASP2
	SEQ ID NO: 112	DEX0249_112	CLASP5 CLASP1
	SEQ ID NO: 113	DEX0249_113	CLASP5 CLASP1
	SEQ ID NO: 114	DEX0249_114	CLASP5 CLASP1
	SEQ ID NO: 115	DEX0249_115	CLASP2
15	SEQ ID NO: 116	DEX0249_116	CLASP2
	SEQ ID NO: 117	DEX0249_117	CLASP2
	SEQ ID NO: 118	DEX0249_118	CLASP2
	SEQ ID NO: 119	DEX0249_119	CLASP2
	SEQ ID NO: 120	DEX0249_120	CLASP2
20	SEQ ID NO: 121	DEX0249_121	CLASP2
	SEQ ID NO: 122	DEX0249_122	CLASP1
	SEQ ID NO: 123	DEX0249_123	CLASP1
	SEQ ID NO: 124	DEX0249_124	CLASP5 CLASP1
	SEQ ID NO: 125	DEX0249_125	CLASP5 CLASP1
25	SEQ ID NO: 126	DEX0249_126	CLASP2
	SEQ ID NO: 127	DEX0249_127	CLASP2
	SEQ ID NO: 128	DEX0249_128	CLASP2
	SEQ ID NO: 129	DEX0249_129	CLASP5 CLASP1
	SEQ ID NO: 130	DEX0249_130	CLASP5 CLASP1
30	SEQ ID NO: 131	DEX0249_131	CLASP5 CLASP1
	SEQ ID NO: 132	DEX0249_132	CLASP2
	SEQ ID NO: 133	DEX0249_133	CLASP2
	SEQ ID NO: 134	DEX0249_134	CLASP2
	SEQ ID NO: 135	DEX0249_135	CLASP1
35	SEQ ID NO: 136	DEX0249_136	CLASP1
	SEQ ID NO: 137	DEX0249_137	CLASP1
	SEQ ID NO: 138	DEX0249_138	CLASP2 CLASP1
	SEQ ID NO: 139	DEX0249_139	CLASP2
	SEQ ID NO: 140	DEX0249_140	CLASP2
40	SEQ ID NO: 141	DEX0249_141	CLASP2
	SEQ ID NO: 142	DEX0249_142	CLASP2
	SEQ ID NO: 143	DEX0249_143	CLASP2
	SEQ ID NO: 144	DEX0249_144	CLASP2
	SEQ ID NO: 145	DEX0249_145	CLASP2
45	SEQ ID NO: 146	DEX0249_146	CLASP2
	SEQ ID NO: 147	DEX0249_147	CLASP2
	SEQ ID NO: 148	DEX0249_148	CLASP2
	SEQ ID NO: 149	DEX0249_149	CLASP2

-121-

	SEQ ID NO: 150	DEX0249_150	CLASP2
	SEQ ID NO: 151	DEX0249_151	CLASP2
	SEQ ID NO: 152	DEX0249_152	CLASP2
	SEQ ID NO: 153	DEX0249_153	CLASP2
5	SEQ ID NO: 154	DEX0249_154	CLASP2
	SEQ ID NO: 155	DEX0249_155	CLASP2
	SEQ ID NO: 156	DEX0249_156	CLASP2
	SEQ ID NO: 158	DEX0249_158	CLASP2
10	SEQ ID NO: 159	DEX0249_159	CLASP2

### Example 2: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'-3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the BSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to normal thymus (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the BSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to normal thymus (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, the BSNA's that show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples.

Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 159 being a diagnostic marker for cancer.

#### DEX0093\_11 (mam029-sqmam036)

Sequence	Sequence ID No
Dex0093_11 (mam029-sqmam036)	DEX0249_15 (SEQ ID NO:15)
	DEX0249_16 (SEQ ID NO:16)

Semi-quantitative PCR was done using the following primers:

Primer	DexSeqID	From	To	Primer Length
sqmam036F	DEX0249_16	395	417	23
sqmam036F	DEX0249_15	50	72	23
sqmam036R	DEX0249_16	684	661	24
sqmam036R	DEX0249_15	339	316	24

Data from the semiQ-PCR experiment showed that sqmam036 was expressed in normal breast in higher levels compared with other 11 normal tissue samples tested, and

-123-

expressed in breast carcinoma while absent in other 11 cancer tissue types (bladder, colon, kidney, lung, ovary, liver, pancreas, prostate, stomach, testis and uterus). Sqmam123 was advanced to quantitative PCR and named mam029.

5 Quantitative PCR was done using the following primers:

Primer	DexSeqID	From	To	Primer Length
mam029F	DEX0249 16	859	876	18
mam029F	DEX0249 15	514	531	18
mam029R	DEX0249 16	992	970	23
mam029R	DEX0249 15	647	625	23
mam029 probe	DEX0249 16	961	934	28
mam029 probe	DEX0249 15	616	589	28

The relative levels of expression of mam029 in 36 normal samples from 25 different tissues were measured. All the values are compared to normal testis (calibrator). These RNA samples are commercially available pools, originated by pooling  
 10 samples of a particular tissue from different individuals; except for the blood samples that they are normal samples from a single individual.

Tissue	Normal
Adrenal Gland	0.0018
Bladder	0.0000
Brain	0.0013
Cervix	0.0154
Colon	0.0000
Endometrium	0.0000
Esophagus	0.0016
Heart	0.0000
Kidney	0.0000
Liver	0.0000
Lung	0.0004
Mammary gland	3.1167
Muscle	0.0041
Ovary	0.00
Pancreas	0.00
Prostate	0.00
Rectum	0.00
Small Intestine	0.00
Spleen	0.00
Stomach	0.00
Testis	1.00
Thymus	0.00
Trachea	0.00
Uterus	0.00



Blood	B1	0.49
Blood	B10	0.00
Blood	B11	0.00
Blood	B12	0.00
Blood	B13	0.00
Blood	B14	0.00
Blood	B15	0.00
Blood	B4	0.00
Blood	B5	0.00
Blood	B6	0.00
Blood	B7	0.00
Blood	B8	0.00

The relative levels of expression in Table 1 show that mam029 mRNA is highly expressed in the pool of normal mammary gland compared with the other normal tissue analyzed.

- 5        The absolute numbers in Table 1 were obtained analyzing pools of samples of a particular tissue from different individuals. They cannot be compared to the absolute numbers originated from RNA obtained from tissue samples of a single individual in Table 2.

- 10        The relative levels of expression of mam029 in 48 pairs of matching samples were measured. All the values are compared to normal testis (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. In addition, 1 unmatched cancer sample (from ovary) and 1 unmatched normal sample (from ovary) were also tested.

15

Sample ID	Tissue	Cancer	Normal Adjacent Tissue	Normal
Mam497M	mammary gland 1	34.66	7.36	
Mam173M	mammary gland 2	0.26	25.28	
Mam726M	mammary gland 3	26.45	7.44	
MamS516	mammary gland 4	0.48	0.22	
MamS621	mammary gland 5	7.70	0.47	
MamS079	mammary gland 6	0.92	1.32	
Mam517	mammary gland 7	0.40	0.83	
Mam19DN	mammary gland 8	27.76	16.97	
Mam522	mammary gland 9	13.09	1.39	
MamB011X	mammary gland 10	0.02	4.77	
MamS127	mammary gland 11	0.00	0.63	
Mam51DN	mammary gland 12	1.79	4.69	
Mam220	mammary gland 13	5.37	18.25	
Mam245M	mammary gland 14	0.02	1.72	

-125-

Mam162X	mammary gland 15	5.31	0.15	
MamS123	mammary gland 16	0.06	2.35	
MamS997	mammary gland 17	0.52	1.04	
Mam543M	mammary gland 18	0.00	0.00	
Mam976M	mammary gland 19	14.47	1.64	
Mam76DN	mammary gland 20	28.84	44.63	
Mam699F	mammary gland 21	0.03	0.56	
Mam42DN	mammary gland 22	0.17	2.46	
MamS570	mammary gland 23	0.04	19.23	
MamS918	mammary gland 24	24.50	12.34	
MamS854	mammary gland 25	3.19	0.76	
Mam986	mammary gland 26	9.71	5.78	
MamS967	mammary gland 27	45.10	2.66	
Mam355	mammary gland 28	56.69	0.10	
MamS699	mammary gland 29	0.30	1.89	
MamA06X	mammary gland 30	121.1	1.25	
Bld32XK	bladder 1	0.01	0.00	
Bld66X	bladder 2	0.00	0.00	
BldTR17	bladder 3	0.00	0.00	
Bld46XK	bladder 4	0.00	0.00	
BldTR14	bladder 5	0.00	0.00	
ClnB56	colon 1	0.00	0.00	
ClnDC63	colon 2	0.007366	0.00	
CvxKS52	cervix 1	0.00	0.00	
CvxNK24	cervix 2	0.00	0.00	
CvxKS83	cervix 3	0.00	0.00	
CvxNK23	cervix 4	0.00	0.00	
Endo10479	endometrium 1	0.00	0.07	
Endo12XA	endometrium 2	0.00	0.00	
Endo5XA	endometrium 3	0.00	0.00	
Endo65RA	endometrium 4	0.00	0.00	
Endo28XA	endometrium 5	0.00	0.00	
Endo3AX	endometrium 6	0.52	0.00	
Kid98XD	kidney 1	0.00	0.00	
Kid6XD	kidney 2	0.00	0.00	
Kid710K	kidney 3	0.00	0.00	
Liv175L	liver 1	0.00	0.00	
Liv187L	liver 2	0.00	0.00	
Liv15XA	liver 3	0.07	0.00	
Lng47XQ	lung 1	0.08	0.00	
LngAC88	lung 2	0.00	0.00	
LngAC90	lung 3	0.00	0.00	
LngSQ80	lung 4	0.00	0.00	
Ovr1118	ovary 1	0.00		
Ovr32RA	ovary 2			0.00
Pan77X	pancreas 1	0.00	0.00	
Pan82XP	pancreas 2	0.01	0.00	
Pro109XB	prostate	0.00	0.08	
Skn248S	skin 1	0.00	0.04	
Skn287S	skin 2	0.00	0.00	
SmIntH89	small intestine 1	0.00	0.00	
SmInt21XA	small intestine 2	0.01	0.00	
Sto115S	stomach 1	0.00	0.00	

-126-

Sto15S	stomach 2	0.00	0.00	
StoMT54	stomach 3	0.00	0.00	
Thr590D	thymus 1	0.00	0.00	
Tst647T	testis 1	0.00	0.09	
Utr141XO	uterus 1	0.01	0.00	
Utr23XU	uterus 2	0.00	0.00	
Utr85XU	uterus 3	0.05	0.00	
Utr135XO	uterus 4	0.00	0.00	

0.00= Negative

The table above represents 148 samples in 17 different tissues. Table 1 and Table 2 represent a combined total of 184 samples in 27 human tissue types.

- 5 Comparisons of the level of mRNA expression in breast cancer samples and the normal adjacent tissue from the same individuals are shown in Table 2. Mam029 is expressed at higher levels in 14 of 30 (47%) cancer samples (mammary gland 1, 3-5, 8, 9, 15, 19, 24-28, 30) compared to normal adjacent tissue.

- Mam029 is highly specific for breast, showing no expression in most of the  
10 samples analyzed for other cancers (specificity: 98%; specificity was calculated as the percentage of samples other than breast with level of expression below of 1/10 of the median for breast cancer samples).

Northern blot analysis for mam029 revealed a 1.2kb transcript.

#### 15 DEX0093\_35 Sqmam046

Sequence

Dex0093\_35 (sqmam046)

Sequence ID #

DEX0249\_50 (SEQ ID NO:50)

- 20 Semi-quantitative PCR was done using the following primers:

Primer	DexSeqID	From	To	Primer Length
sqmam046F	DEX0249_50	3	24	22
sqmam046R	DEX0249_50	380	361	20

- The relative levels of expression of sqmam046 in 12 normal samples from 12 different tissues are shown below. These RNA samples are from single individual or are  
25 commercially available pools, originated by pooling samples of a particular tissue from different individuals.. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

-127-

TISSUE	NORMAL
Breast	1
Colon	10
Endometrium	10
Kidney	0
Liver	0
Lung	1
Ovary	1
Prostate	100
Small Intestine	1
Stomach	100
Testis	1
Uterus	0

Relative levels of expression in the table above show that normal prostate and stomach exhibit the highest expression of sqmam046, followed by colon and

5 endometrium.

The relative levels of expression of sqmam046 in 12 cancer samples from 12 different tissues are shown below. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression.

10 A positive reaction in the most dilute sample indicates the highest relative expression value.

TISSUE	CANCER
bladder	1
breast	100
colon	1
kidney	0
liver	0
lung	0
ovary	0
pancreas	0
prostate	10
stomach	0
testes	0
Uterus	1

Relative levels of expression in the table above shows that sqmam046 is highly

15 expressed in breast carcinoma when compared with the other cancer samples tested.

The relative levels of expression of sqmam046 in 6 mammary gland cancer matching samples are shown below. A matching pair is formed by mRNA from the

cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

SAMPLE ID	TISSUE	CANCER	NORMAL ADJACENT TISSUE
S99522A/B	mammary gland 1	10	0
4005724A2/B3	mammary gland 2	1	1
4005599A4/B2	mammary gland 3	1	0
4005629A2/B2	mammary gland 4	10	1
S9822245A/B	mammary gland 5	100	10
S9819997A/B	mammary gland 6	1	1

Relative levels of expression in the Table above shows that sqmam046 is upregulated in 4 out of 6 (67%) of the matching samples analyzed. Experiments are underway to design and test primers and probe for quantitative PCR.

#### DEX0093\_46

#### 15 Sqmam050

Sequence

Dex0093\_46 (sqmam050)

Sequence ID #

DEX0249\_69 (SEQ ID NO:69)

DEX0249\_70 (SEQ ID NO:70)

20

Semi-quantitative PCR was done using the following primers:

Primer	DexSeqID	From	To	Primer Length
sqmam050F	DEX0249_70	68	92	25
sqmam050F	DEX0249_69	68	92	25
sqmam050R	DEX0249_70	523	502	22
sqmam050R	DEX0249_69	523	502	22

The relative levels of expression of sqmam050 in 12 normal samples from 12 different tissues are shown below. These RNA samples are from single individual or are commercially available pools, originated by pooling samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative

expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

TISSUE	NORMAL
Breast	0
Colon	0
Endometrium	0
Kidney	0
Liver	0
Lung	0
Ovary	0
Prostate	0
Small Intestine	0
Stomach	0
Testis	0
Uterus	0

- 5 Relative levels of expression in Table 1 show that none of the normal tissues tested show expression of sqmam050.

The relative levels of expression of sqmam050 in 12 cancer samples from 12 different tissues are shown below. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate.

- 10 Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

TISSUE	CANCER
bladder	1
breast	0
colon	1
kidney	1
liver	0
lung	1
ovary	0
pancreas	1
prostate	10
stomach	0
testes	1
uterus	10

- 15 Relative levels of expression in Table 2 show that sqmam050 is expressed in the cancer samples tested, with the highest expression in prostate and uterus.

The relative levels of expression of sqmam050 in 6 mammary gland cancer matching samples are show below. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

- 5 Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

SAMPLE ID	TISSUE	CANCER	NORMAL ADJACENT TISSUE
S99522A/B	mammary gland 1	10	1
4005724A2/B3	mammary gland 2	1	0
4005599A4/B2	mammary gland 3	1	0
4005629A2/B2	mammary gland 4	1	1
S9822245A/B	mammary gland 5	0	0
S9819997A/B	mammary gland 6	10	1

10

Relative levels of expression in Table 2 shows that sqmam050 is upregulated in 4 out of 6 (67%) of the matching samples analyzed.

Experiments are underway to design and test primers and probe for quantitative PCR.

15

DEX0093\_65  
Sqmam054

Sequence Sequence ID #  
20 Dex0093\_65 (sqmam054) DEX0249\_93 (SEQ ID NO: 93)  
DEX0249\_94 (SEQ ID NO:94)

Semi-quantitative PCR was done using the following primers:

Primer	DexSeqID	From	To	Primer Length
sqmam054F	DEX0249_94	115	132	18
sqmam054F	DEX0249_93	117	134	18
sqmam054R	DEX0249_94	251	234	18
sqmam054R	DEX0249_93	253	236	18

25

Table 1. The absolute numbers are relative levels of expression of sqmam054 in 12 normal samples from 12 different tissues. These RNA samples are from single individual or are commercially available pools, originated by pooling samples of a

-131-

particular tissue from different individuals.. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

TISSUE	NORMAL
Breast	0
Colon	0
Endometrium	0
Kidney	0
Liver	0
Lung	0
Ovary	0
Prostate	0
Small Intestine	0
Stomach	1
Testis	0
Uterus	1

Relative levels of expression in Table 1 show expression of sqmam054 in stomach and uterus normal tissues.

**Table 2.** The relative levels of expression of sqmam054 in 12 cancer samples from 12 different tissues are shown below. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

TISSUE	CANCER
bladder	1
breast	1
colon	0
kidney	1
liver	0
lung	1
ovary	0
pancreas	1
prostate	1
stomach	10
Testes	1
Uterus	1



Relative levels of expression in Table 2 show that sqmam054 is expressed in some of the cancer samples tested. The highest level of expression in stomach cancer tissue.

**Table 3.** The absolute numbers are relative levels of expression of sqmam054 in 6 mammary gland cancer matching samples. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

SAMPLE ID	TISSUE	CANCER	NORMAL ADJACENT TISSUE
S99522A/B	mammary gland 1	1	0
4005724A2/B3	mammary gland 2	0	0
4005599A4/B2	mammary gland 3	1	0
4005629A2/B2	mammary gland 4	1	0
S9822245A/B	mammary gland 5	0	10
S9819997A/B	mammary gland 6	0	0

Relative levels of expression in Table 2 shows that sqmam054 is upregulated in 3 out of 6 (50%) of the matching samples analyzed, and two matching samples do not express sqmam054 in any of both samples: cancer and NAT.

Experiments are underway to design and test primers and probe for quantitative PCR.

## 20 Example 2B: Custom Microarray Experiment

Custom oligonucleotide microarrays were provided by Agilent Technologies, Inc. (Palo Alto, CA). The microarrays were fabricated by Agilent using their technology for the *in-situ* synthesis of 60mer oligonucleotides (Hughes, et al. 2001, Nature Biotechnology 19:342-347). The 60mer microarray probes were designed by Agilent, from gene sequences provided by diaDexus, using Agilent proprietary algorithms. Whenever possible two different 60mers were designed for each gene of interest.

All microarray experiments were two-color experiments and were preformed using Agilent-recommended protocols and reagents. Briefly, each microarray was

- hybridized with cRNAs synthesized from polyA+ RNA, isolated from cancer and normal tissues, labeled with fluorescent dyes Cyanine3 and Cyanine5 (NEN Life Science Products, Inc., Boston, MA) using a linear amplification method (Agilent). In each experiment the experimental sample was polyA+ RNA isolated from cancer tissue from a
- 5 single individual and the reference sample was a pool of polyA+ RNA isolated from normal tissues of the same organ as the cancerous tissue (*i.e.* normal breast tissue in experiments with breast cancer samples). Hybridizations were carried out at 60°C, overnight using Agilent *in-situ* hybridization buffer. Following washing, arrays were scanned with a GenePix 4000B Microarray Scanner (Axon Instruments, Inc., Union City,
- 10 CA). The resulting images were analyzed with GenePix Pro 3.0 Microarray Acquisition and Analysis Software (Axon). A total of 20 experiments comparing the expression patterns of breast cancer derived polyA+ RNA (6 stage 1 cancers, 12 stage 2 cancers, 2 stage 3 cancers) to polyA+ RNA isolated from a pool of 10 normal breast tissues were analyzed.
- 15 Data normalization and expression profiling were done with Expressionist software from GeneData Inc. (Daly City, CA/Basel, Switzerland). Gene expression analysis was performed using only experiments that meet certain quality criteria. The quality criteria that experiments must meet are a combination of evaluations performed by the Expressionist software and evaluations performed manually using raw and
- 20 normalized data. To evaluate raw data quality, detection limits (the mean signal for a replicated negative control + 2 Standard Deviations (SD)) for each channel were calculated. The detection limit is a measure of non-specific hybridization. Arrays with poor detection limits were not analyzed and the experiments were repeated. To evaluate normalized data quality, positive control elements included in the array were utilized.
- 25 These array features should have a mean ratio of 1 (no differential expression). If these features have a mean ratio of greater than 1.5-fold up or down, the experiments were not analyzed further and were repeated. In addition to traditional scatter plots demonstrating the distribution of signal in each experiment, the Expressionist software also has minimum thresholding criteria that employ user defined parameters to identify quality
- 30 data. Only those features that meet the threshold criteria were included in the filtering and analyses carried out by Expressionist. The thresholding settings employed require a minimum area percentage of 60% [(% pixels > background + 2SD)-(% pixels saturated)],

and a minimum signal to noise ratio of 2.0 in both channels. By these criteria, very low expressors and saturated features were not included in analysis.

Relative expression data was collected from Expressionist based on filtering and clustering analyses. Up- and down- regulated genes were identified using criteria for percentage of valid values obtained, and the percentage of experiments in which the gene is up- or down-regulated. These criteria were set independently for each data set, depending on the size and the nature of the data set. The results for the statistically significant upregulated and downregulated genes are shown in Table 1. The first three columns of the table contain information about the sequence itself (Oligo ID, Parent ID, and Patent#), the next 3 columns show the results obtained. '%valid' indicates the percentage of 20 unique experiments total in which a valid expression value was obtained, '%up' indicates the percentage of 20 experiments in which up-regulation of at least 2.5-fold was observed, and '%down' indicates the percentage of the 20 experiments in which down-regulation of at least 2.5-fold was observed. The last column in Table 1 describes the location of the microarray probe (oligo) relative to the parent sequence. Additional sequences were examined but the data were inconclusive.

Sensitivity data for DEX0093 series microarray features.

OligoID	Parent ID	Patent #	Sensitivity of up and down regulation			Oligo Seq location
			% valid	%up	%down	
20173	4795	DEX0093_86 DEX0249_124	65%	40%	5%	653-712

### Example 3: Protein Expression

The BSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the BSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the BSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH<sub>2</sub>-terminus of the coding sequence of BSNA, and six histidines, flanking the COOH-terminus of the coding sequence of BSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of BSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickle chelating resin. The column was packed and washed with five column  
5 volumes of wash buffer. BSP was eluted stepwise with various concentration imidazole buffers.

#### **Example 4: Protein Fusions**

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also  
10 should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a  
15 polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the  
20 vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

#### **Example 5: Production of an Antibody from a Polypeptide**

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such  
25 cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any  
30 suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from

the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

- The hybridoma cells obtained through such a selection are then assayed to
- 5 identify clones which secrete antibodies capable of binding the polypeptide.
- Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method,
- 10 protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be
  - 15 used to immunize an animal to induce formation of further protein-specific antibodies.
- Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by reference).

- Examples of post-translational modifications (PTMs) of the BSPs of this
- 20 invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. Using the ProSite database (Bairoch *et al.*, Nucleic Acids Res. 25(1):217-221 (1997), the contents of which are incorporated by reference), the following PTMs were predicted for the BSPs of the invention ([http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=npsa\\_prosite.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_prosite.html)
  - 25 most recently accessed October 23, 2001). For full definitions of the PTMs see <http://www.expasy.org/cgi-bin/prosite-list.pl> most recently accessed October 23, 2001.

- |                |                                                 |
|----------------|-------------------------------------------------|
| DEX0249_160    | Asn_Glycosylation 2-5; Ck2_Phospho_Site 4-7;    |
| DEX0249_161    | Pkc_Phospho_Site 11-13;                         |
| DEX0249_162    | Myristyl 2-7;                                   |
| 30 DEX0249_163 | Ck2_Phospho_Site 2-5;42-45;                     |
| DEX0249_164    | Myristyl 16-21; Pkc_Phospho_Site 20-22;         |
| DEX0249_165    | Ck2_Phospho_Site 26-29;                         |
| DEX0249_166    | Asn_Glycosylation 4-7;                          |
| DEX0249_167    | Asn_Glycosylation 9-12;                         |
| 35 DEX0249_168 | Ck2_Phospho_Site 48-51; Pkc_Phospho_Site 59-61; |

	DEX0249_169	Ck2_Phospho_Site 81-84; Myristyl 17-22;26-31;42-47; Pkc_Phospho_Site 10-12;83-85;
	DEX0249_170	Ck2_Phospho_Site 3-6; Pkc_Phospho_Site 5-7;
	DEX0249_171	Camp_Phospho_Site 69-72; Ck2_Phospho_Site 20-23; Myristyl
5		13-18; Pkc_Phospho_Site 20-22;54-56;63-65;68-70;
	DEX0249_172	Pkc_Phospho_Site 25-27;
	DEX0249_173	Glycosaminoglycan 25-28; Pkc_Phospho_Site 25-27;42-44;
	DEX0249_174	Myristyl 20-25; Pkc_Phospho_Site 6-8;
	DEX0249_175	Asn_Glycosylation 2-5; Myristyl 16-21;
10	DEX0249_176	Pkc_Phospho_Site 8-10;
	DEX0249_177	Ck2_Phospho_Site 4-7;
	DEX0249_178	Pkc_Phospho_Site 8-10;
	DEX0249_179	Myristyl 34-39; Pkc_Phospho_Site 51-53;
	DEX0249_180	Ck2_Phospho_Site 19-22;21-24; Myristyl 13-18;38-43;
15	DEX0249_182	Ck2_Phospho_Site 3-6;39-42; Pkc_Phospho_Site 3-5;39-41;
	DEX0249_183	Ck2_Phospho_Site 39-42; Myristyl 17-22; Pkc_Phospho_Site 53-55;
	DEX0249_184	Myristyl 18-23;
	DEX0249_187	Myristyl 39-44;
20	DEX0249_188	Myristyl 13-18;52-57;68-73; Pkc_Phospho_Site 32-34;56-58;64-66;65-67;
	DEX0249_189	Myristyl 17-22; Pkc_Phospho_Site 22-24;
	DEX0249_190	Amidation 6-9; Ck2_Phospho_Site 27-30; Tyr_Phospho_Site 8-16;9-16;
25	DEX0249_192	Ck2_Phospho_Site 23-26; Pkc_Phospho_Site 23-25;
	DEX0249_193	Amidation 217-220; Camp_Phospho_Site 84-87;
		Ck2_Phospho_Site 20-23;22-25;31-34; Myristyl 18-23;42-47;57-62;58-63;73-78;135-140;175-180;201-206;204-209;285-290;
		Pkc_Phospho_Site 139-141;286-288;289-291;
30	DEX0249_194	Pkc_Phospho_Site 26-28;
	DEX0249_195	Myristyl 16-21; Pkc_Phospho_Site 36-38;70-72;
	DEX0249_196	Camp_Phospho_Site 74-77; Ck2_Phospho_Site 79-82; Myristyl
		12-17; Pkc_Phospho_Site 16-18;30-32;73-75;79-81;
	DEX0249_197	Asn_Glycosylation 203-206;243-246;485-488;548-551;594-597;
35		Ck2_Phospho_Site 176-179;337-340;345-348;389-392;398-401;407-410;431-434;452-455;465-468;476-479;503-506;550-553;563-566; Myristyl 520-525;538-543;558-563;
		Pkc_Phospho_Site 45-47;176-178;195-197;205-207;217-219;221-223;232-234;312-314;315-317;320-322;362-364;365-367;413-415;431-433;530-532;614-616;633-635; Prokar_Lipoprotein 133-143;
40		
	DEX0249_198	Amidation 7-10; Pkc_Phospho_Site 5-7;
	DEX0249_199	Pkc_Phospho_Site 3-5;
	DEX0249_200	Myristyl 32-37;
45	DEX0249_202	Ck2_Phospho_Site 49-52;53-56;
	DEX0249_204	Asn_Glycosylation 146-149; Ck2_Phospho_Site 19-22;43-46;73-76;120-123; Pkc_Phospho_Site 50-52;127-129;131-133;

	DEX0249_205	Ck2_Phospho_Site 3-6;85-88; Myristyl 108-113; Pkc_Phospho_Site 44-46;93-95;
	DEX0249_206	Pkc_Phospho_Site 12-14;20-22;23-25;
	DEX0249_207	Ck2_Phospho_Site 103-106; Myristyl 46-51;71-76;
5		Pkc_Phospho_Site 39-41;121-123;
	DEX0249_208	Asn_Glycosylation 37-40;
	DEX0249_209	Ck2_Phospho_Site 7-10;23-26;28-31; Pkc_Phospho_Site 22-24;
	DEX0249_210	Myristyl 55-60; Pkc_Phospho_Site 21-23; Tyr_Phospho_Site 50-57;
10	DEX0249_212	Pkc_Phospho_Site 23-25;
	DEX0249_213	Pkc_Phospho_Site 22-24;
	DEX0249_215	Asn_Glycosylation 80-83;127-130; Ck2_Phospho_Site 18-21; Myristyl 2-7;56-61;71-76;86-91;116-121;120-125; Pkc_Phospho_Site 75-77; Prokar_Lipoprotein 9-19;
15	DEX0249_216	Ck2_Phospho_Site 41-44;
	DEX0249_217	Pkc_Phospho_Site 57-59;
	DEX0249_218	Leucine_Zipper 22-43; Pkc_Phospho_Site 42-44;59-61; Prokar_Lipoprotein 50-60; Tyr_Phospho_Site 61-68;
	DEX0249_219	Pkc_Phospho_Site 8-10;
20	DEX0249_220	Asn_Glycosylation 18-21; Pkc_Phospho_Site 20-22;
	DEX0249_222	Ck2_Phospho_Site 9-12;22-25; Pkc_Phospho_Site 19-21; Tyr_Phospho_Site 21-28;
	DEX0249_223	Ck2_Phospho_Site 4-7; Myristyl 12-17;21-26; Pkc_Phospho_Site 4-6;
25	DEX0249_224	Myristyl 31-36; Pkc_Phospho_Site 42-44;
	DEX0249_225	Pkc_Phospho_Site 19-21;
	DEX0249_227	Ck2_Phospho_Site 29-32; Myristyl 40-45;
	DEX0249_228	Ck2_Phospho_Site 8-11;14-17; Pkc_Phospho_Site 22-24;76-78;
	DEX0249_229	Ck2_Phospho_Site 68-71; Myristyl 6-11;50-55;
30		Pkc_Phospho_Site 62-64;
	DEX0249_230	Myristyl 5-10; Pkc_Phospho_Site 21-23;
	DEX0249_231	Asn_Glycosylation 39-42; Pkc_Phospho_Site 43-45;
	DEX0249_233	Myristyl 25-30;
35	DEX0249_234	Amidation 145-148; Ck2_Phospho_Site 29-32;30-33;56-59;96-99;98-101; Myristyl 85-90;92-97;94-99; Pkc_Phospho_Site 56-58;63-65;103-105;
	DEX0249_235	Asn_Glycosylation 56-59;171-174;399-402; Camp_Phospho_Site 51-54;873-876; Ck2_Phospho_Site 13-16;123-126;182-185;222-225;266-269;320-323;649-652;682-685;795-798;832-835;845-848;892-895; Myristyl 108-113;315-320;468-473;499-504;581-586; Pkc_Phospho_Site 60-62;101-103;146-148;163-165;166-168;222-224;279-281;289-291;336-338;337-339;391-393;574-576;607-609;655-657;673-675;784-786;809-811;832-834;856-858;920-922;932-934; Tyr_Phospho_Site 431-438;
40		
	DEX0249_236	Pkc_Phospho_Site 47-49;
	DEX0249_237	Ck2_Phospho_Site 23-26;31-34;
	DEX0249_238	Myristyl 20-25;
45	DEX0249_239	Camp_Phospho_Site 11-14; Pkc_Phospho_Site 9-11;

	DEX0249_240	Camp_Phospho_Site 34-37; Helix_Loop_Helix 13-28;
	DEX0249_241	Ck2_Phospho_Site 20-23;34-37; Myristyl 3-8;
	DEX0249_243	Asn_Glycosylation 26-29;33-36; Ck2_Phospho_Site 35-38;
		Pkc_Phospho_Site 16-18;
5	DEX0249_244	Myristyl 13-18; Pkc_Phospho_Site 52-54;
	DEX0249_245	Ck2_Phospho_Site 17-20; Pkc_Phospho_Site 11-13;17-19;
	DEX0249_246	Amidation 10-13; Asn_Glycosylation 24-27;
	DEX0249_249	Asn_Glycosylation 19-22;
	DEX0249_250	Ck2_Phospho_Site 153-156; Myristyl 110-115;124-129;125-
10		130;129-134;140-145; Pkc_Phospho_Site 3-5;20-22;
	DEX0249_251	Asn_Glycosylation 40-43; Myristyl 22-27;
	DEX0249_252	Ck2_Phospho_Site 17-20;21-24;152-155; Myristyl 22-27;28-
		33;100-105;115-120; Pkc_Phospho_Site 17-19;152-154;156-158;
	DEX0249_253	Amidation 26-29; Asn_Glycosylation 51-54; Ck2_Phospho_Site
15		45-48; Myristyl 50-55;
	DEX0249_255	Asn_Glycosylation 6-9;31-34; Myristyl 58-63; Pkc_Phospho_Site
		38-40;
	DEX0249_256	Asn_Glycosylation 6-9;
	DEX0249_257	Ck2_Phospho_Site 2-5;9-12;
20	DEX0249_259	Ck2_Phospho_Site 4-7;
	DEX0249_260	Asn_Glycosylation 208-211;294-297;601-604;
		Camp_Phospho_Site 154-157;669-672; Ck2_Phospho_Site 247-
		250;338-341;352-355;513-516;519-522;527-530;551-554;556-
		559;606-609;637-640;683-686;739-742;747-750;756-759;795-
25		798;817-820;871-874;885-888;898-901;904-907;912-915;
		Leucine_Zipper 284-305;408-429;415-436;422-443;429-450;436-
		457;443-464;450-471;555-576;562-583;569-590;576-597;969-
		990; Myristyl 189-194;386-391;736-741;942-947;
		Pkc_Phospho_Site 38-40;210-212;270-272;371-373;519-521;561-
30		563;739-741;792-794;898-900;912-914;
	DEX0249_261	Asn_Glycosylation 9-12; Ck2_Phospho_Site 11-14;
		Pkc_Phospho_Site 34-36;
	DEX0249_263	Ck2_Phospho_Site 62-65; Myristyl 22-27;
	DEX0249_264	Pkc_Phospho_Site 8-10;
35	DEX0249_265	Amidation 381-384;452-455; Asn_Glycosylation 219-222;427-
		430; Ck2_Phospho_Site 58-61;95-98;111-114;378-381;415-
		418;441-444;465-468; Dead_Atp_Helicase 40-48; Myristyl 26-
		31;70-75;100-105;146-151;191-196;448-453; Pkc_Phospho_Site
		95-97;230-232;452-454;456-458;
40	DEX0249_266	Pkc_Phospho_Site 11-13;18-20;
	DEX0249_268	Amidation 26-29; Camp_Phospho_Site 28-31; Pkc_Phospho_Site
		14-16;40-42; Rgd 3-5;
	DEX0249_269	Pkc_Phospho_Site 13-15;
	DEX0249_271	Asn_Glycosylation 39-42;45-48; Camp_Phospho_Site 9-12;
45	DEX0249_272	Ck2_Phospho_Site 11-14;
	DEX0249_273	Ck2_Phospho_Site 6-9;
	DEX0249_274	Ck2_Phospho_Site 24-27; Myristyl 60-65;



	DEX0249_275	Amidation 201-204; Asn_Glycosylation 12-15; Camp_Phospho_Site 128-131;193-196; Ck2_Phospho_Site 208-211;346-349;370-373; Leucine_Zipper 413-434;420-441; Myristyl 221-226;237-242; Pkc_Phospho_Site 13-15;126-128;127-129;143-145;176-178;186-188;214-216;355-357;
5	DEX0249_277	Asn_Glycosylation 98-101;149-152; Myristyl 147-152; Pkc_Phospho_Site 151-153;166-168; Tyr_Phospho_Site 25-33;26-33;72-79;75-82;
	DEX0249_278	Asn_Glycosylation 141-144; Ck2_Phospho_Site 21-24; Tyr_Phospho_Site 68-76;69-76;115-122;118-125;
10	DEX0249_279	Pkc_Phospho_Site 13-15;
	DEX0249_281	Ck2_Phospho_Site 49-52;80-83;
	DEX0249_282	Amidation 77-80;217-220;389-392;831-834; Asn_Glycosylation 83-86;663-666; Camp_Phospho_Site 277-280;596-599;
15		Ck2_Phospho_Site 16-19;29-32;65-68;204-207;213-216;244-247;271-274;576-579;610-613;611-614;615-618;621-624;705-708;775-778;931-934; Glycosaminoglycan 22-25;86-89;88-91; Myristyl 73-78;87-92;89-94;91-96;94-99;148-153;214-219;268-273;284-289;294-299;297-302;347-352;380-385;429-434;445-450;451-456;454-459;455-460;464-469;469-474;500-505;517-522;530-535;587-592;681-686;734-739;739-744;802-807;853-858;870-875;922-927;971-976; Pkc_Phospho_Site 77-79;166-168;217-219;264-266;327-329;440-442;472-474;534-536;588-590;699-701;716-718;926-928;940-942;956-958;972-974;
20		

#### **Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide**

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1 through 159. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al.*, *Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is

cloned into T-tailed vectors as described in Holton *et al.*, *Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are  
5 nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Mannheim), and FISH is performed as described in Johnson *et al.*, *Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium  
10 iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical  
15 Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

**Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological  
20 Sample**

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 µg/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells  
25 are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 µl of specific antibody-alkaline phosphatase conjugate,  
30 at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 µl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl

phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

#### **Example 8: Formulating a Polypeptide**

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 ,  $\mu\text{g/kg/day}$  to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1  $\mu\text{g/kg/hour}$  to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of

administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable  
5 polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl  
10 acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP  
15 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

20 For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

25 For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more  
30 preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-

aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of

pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

**Example 9: Method of Treating Decreased Levels of the Polypeptide**

5 It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a  
10 pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100  $\mu\text{g/kg}$  of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on  
15 administration and formulation, are provided above.

**Example 10: Method of Treating Increased Levels of the Polypeptide**

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

20 For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0  $\text{mg/kg day}$  for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

25 **Example 11: Method of Treatment Using Gene Therapy**

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a  
30 tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of

the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days.

- 5 After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using  
10 glass beads.

- The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear  
15 backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

- 20 The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are  
25 now referred to as producer cells).

- Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is  
30 removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

- 5       The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

**Example 12: Method of Treatment Using Gene Therapy-*In Vivo***

- Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the  
10 introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

- The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known  
15 in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) Cardiovasc. Res. 35 (3): 470-479, Chao J et al. (1997) Pharmacol. Res. 35 (6): 517-522, Wolff J. A. (1997) Neuromuscul. Disord. 7 (5): 314-318, Schwartz B. et al. (1996) Gene Ther. 3 (5): 405-411, Tsurumi Y. et al. (1996) Circulation 94 (12): 3281-3290 (incorporated herein by reference).

- 20       The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

- The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free  
25 from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85  
30 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain



sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have  
5 shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach,  
10 intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by  
15 the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or  
20 less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05  $\mu\text{g/kg}$  body weight to about 50 mg/kg body  
25 weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and

particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is  
5 determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

10 Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about  
15 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for  
20 protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper  
25 dosages and other treatment parameters in humans and other animals using naked DNA.

### **Example 13: Transgenic Animals**

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea  
30 pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

- Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver et al.,
- 5 Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., *Biotechnology* (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, *Mol Cell.*
- 10 *Biol.* 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., *Science* 259: 1745 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., *Cell* 57: 717-723 (1989); etc. For a review of such techniques, see Gordon, "Transgenic Animals," *Intl.*
- 15 *Rev. Cytol.* 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

- Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence
- 20 (Campell et al., *Nature* 380: 64-66 (1996); Wilmut et al., *Nature* 385: 810813 (1997)).

- The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail
- 25 tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the
- 30 polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed

for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant

expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

#### Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503-512 (1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, *supra*). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or

transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve  
5 expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft;  
10 genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible  
15 cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

20 Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

25 All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments,  
30 which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

-154-

## CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising
  - (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes  
5 an amino acid sequence of SEQ ID NO: 160 through 282;
  - (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID  
NO: 1 through 159;
  - (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid  
molecule of (a) or (b); or
  - 10 (d) a nucleic acid molecule having at least 60% sequence identity to the nucleic  
acid molecule of (a) or (b).
2. The nucleic acid molecule according to claim 1, wherein the nucleic acid  
molecule is a cDNA.  
15
3. The nucleic acid molecule according to claim 1, wherein the nucleic acid  
molecule is genomic DNA.
4. The nucleic acid molecule according to claim 1, wherein the nucleic acid  
20 molecule is a mammalian nucleic acid molecule.
5. The nucleic acid molecule according to claim 4, wherein the nucleic acid  
molecule is a human nucleic acid molecule.
- 25 6. A method for determining the presence of a breast specific nucleic acid  
(BSNA) in a sample, comprising the steps of:
  - (a) contacting the sample with the nucleic acid molecule according to claim 1  
under conditions in which the nucleic acid molecule will selectively hybridize to a breast  
specific nucleic acid; and
  - 30 (b) detecting hybridization of the nucleic acid molecule to a BSNA in the  
sample, wherein the detection of the hybridization indicates the presence of a BSNA in  
the sample.

7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.

5

9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.

10

10. A polypeptide encoded by the nucleic acid molecule according to claim 1.

11. An isolated polypeptide selected from the group consisting of:

- (a) a polypeptide comprising an amino acid sequence with at least 60%  
15 sequence identity to of SEQ ID NO: 160 through 282; or  
(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 159.

12. An antibody or fragment thereof that specifically binds to the polypeptide  
20 according to claim 11.

13. A method for determining the presence of a breast specific protein in a sample, comprising the steps of:

- (a) contacting the sample with the antibody according to claim 12 under  
25 conditions in which the antibody will selectively bind to the breast specific protein; and  
(b) detecting binding of the antibody to a breast specific protein in the sample, wherein the detection of binding indicates the presence of a breast specific protein in the sample.

14. A method for diagnosing and monitoring the presence and metastases of breast cancer in a patient, comprising the steps of:

30



-156-

(a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient; and

(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the breast specific marker in a  
5 normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of breast cancer.

15 16. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient.

17. A method of treating a patient with breast cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein  
15 said administration induces an immune response against the breast cancer cell expressing the nucleic acid molecule or polypeptide.

17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

20

## SEQUENCE LISTING

<110> Sun, Yongming  
 Recipon, Herve  
 Salceda, Susana  
 Liu, Chenghua  
 Turner, Leah  
 DiaDexus, Inc.

<120> Compositions and Methods Relating to Breast Specific  
 Genes and Proteins

<130> DEX-0250

<140>

<141>

<150> 60/243,802

<151> 2000-10-27

<160> 282

<170> PatentIn Ver. 2.1

<210> 1

<211> 207

<212> DNA

<213> Homo sapiens

<400> 1

```
ggaggagacg tgcaatagag atacccaaag aaatacatga ataattcagg agcagatttg 60
catttatcca ctggtactat ttagttgtat ttattagaca gcttcctgcc ctctccaaaa 120
agcttactga gctagtaact atttacaggg ttagccaaag aacacaaaaa agtgatctct 180
attagactgt aagaatatgg tttccct 207
```

<210> 2

<211> 503

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (492)

<223> a, c, g or t

<400> 2

```
ggttgatatt atattacact tctctgaata aaatgccacc ttgatttgct cattctcaat 60
cggcaggagt cttatttcac ctctgtagggt cttactaagt gtgttttagtt ttcaaaagaa 120
accagtgttt ccctagtact taacatgggt ttattacatt tttttgacaa aaattcaaaa 180
ttacatatat ttgttcttc attagcaagt cacacatttt aaaatggcac actcccttcc 240
tcttcgtggt gttatttggt tattttaagg actgtttctg ggtagataag ctctggggtta 300
ttttaaaata cattttacaa tggaaatggc ctggacttga actgaaaagg aaacattatc 360
tgtgttattt cagacacatc agtgatcagt ttagaagata ggatgatttc actaagctta 420
taattcatct taaagctcac ctaaataaaa gtaagtgact aaaatgatct ttttcttcca 480
ggagaggtag gnttaattaa ttg 503
```

<210> 3

<211> 603

<212> DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 3

```

ggttgatatt atattacact tctctgaata aaatgccacc ttgatttgct cattctcaat 60
cggcaggagt cttatttcac ctctgtaggt cttactaagt gtgttttagtt ttcaaaagaa 120
accagtgttt ccctagtact taacatgggt ttattacatt tttttgacaa aaattcaaaa 180
ttacatatat tttgttcttc attagcaagt cacacathtt aaaatggcac actcccttcc 240
tcttcgtgtt gttatttggt tattttaagg actgtttctg ggtagataag ctctgggtta 300
ttttaaaata cattttacaa tggaaatggc ctggacttga actgaaaagg aaacattatc 360
tgtgtttatht cagacacatc agtgatcagt ttagaagata ggatgatttc actaagctta 420
taattcatct taaagctcac ctaaaataaaa gtaagtgtact aaaatgatct ttttcttcca 480
ggagaggtag gattaattaa tggatataat tgtggaatat ttcaggctta tctgattctt 540
ccatcttaaa tctttgagag ttttaaacac attatgtgtc cattactgtt tatatcacat 600
aga

```

603

&lt;210&gt; 4

&lt;211&gt; 534

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (133)

&lt;223&gt; a, c, g or t

&lt;400&gt; 4

```

gaaatgcagc tgacaatgca aggagcaaga ggactcgcac agtgggtgcat ggcagcttgc 60
tgtcattttc tgggcacaga aagtgcgatg gaagggaatg agaaggggaa aaaggaagga 120
tgacaggacg ganggagggga aagaaggaag aggaaaaaag aaaggacagg agaaagggag 180
gaaggettct gccaaaaaat taaaatcaaa tttttgacat tctttttgtt tgcctttttt 240
gaaacaaaat gacacttgcc agacaccagc ttcctggccc atgtcctggg ccttggtatc 300
cagatgacag cagtgtgatc ctgctgtgag ttccttccgt gccttctgat ctgagttcct 360
gaaagcagag agccactcag gaactgctgt ctctcaggcc agctggctgg tgatgggctt 420
ttgaagactc tggctctctc tcctgctgga agagctcccc aggggccacc aggagccagg 480
tgaccgctct cagcctctgt gagctactgg agatcaccag accttcccac atcc

```

534

&lt;210&gt; 5

&lt;211&gt; 928

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (329)

&lt;223&gt; a, c, g or t

&lt;400&gt; 5

```

atgactctta acgagcatgc tgccttcaaa catctgttta acgaagcaca tcttgcaccg 60
cccttaatcc atttaaccct gagtggacac agcacatgtt tcagagagca cagggttggg 120
gcaactgtgc agacactgga gataataaag aaaaacagca ctgggcctat aattgggagt 180
ctagatatac gatatggaaa tgcagctgac aatgcaagga gcaagaggac tcgcacagtg 240
gtgcatggca gcttgctgtc attttctggg cacagaaagt gcgatggaag ggaatgagaa 300
ggggaaaaag gaaggatgac aggacggang gagggaaaga aggaagagga aaaaagaaag 360
gacaggagaa agggaggaag gcttctgcca aaaaattaaa atcaaatttt tgacattctt 420
tttgtttgcc ttttttgaaa caaaatgaca cttgccagac accagcttcc tggcccatgt 480
cctggctcct ggtatccaga tgacagcagt gtgatcctgc tgtgagttcc ttccgtgcct 540
tctgatctga gttcctgaaa gcagagagcc actcaggaaac tgcgtgtctc caggccagct 600
ggctgggtgat gggcttttga agactctggc tctctctcct gctggaagag ctccccagg 660
gccaccagga gccaggtgac cgtctcagc ctctgtgagc tactggagat caccagacct 720

```

```

tcccacatcc cgggcagggtg ccagggcctt taaggaggct ttctgctctg cagggatggt 780
ctgtgggctc cagtattctg gcgagcatca gcttattctc ggcttagtct tcttgctcta 840
tagctcctgc ctctgttttg cttctttttg gtgatgcctg tttcacctta tgaagagggc 900
ctgtaaaatc caagctctgc acaaacc 928

```

<210> 6  
 <211> 368  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (305)  
 <223> a, c, g or t

```

<400> 6
tattaagtga ctttaagtga agttaccac ttctgggtgct agtttgatg ttaatttgtt 60
tttatcctga tacatataat caggtagaat tgggaatttt attcagcctc cgcgtcgggg 120
aacatagaat aactctttat tgactcaaga ctggaatttt cagaaaagtt taaattttta 180
ttttttctag aatttttcaga aaagttaaat agatctgaga cattttttaa tcttttaata 240
tttctagcta tttgtgaata tgcttttctt ctttttaaat aaatataata gctggatgt 300
aaganagcta ttgatgcata tttttatttg gatattctat tgaactctta attggaataa 360
tttccagt 368

```

<210> 7  
 <211> 583  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (61)..(80)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (495)  
 <223> a, c, g or t

```

<400> 7
cacagcagca tctggactgg tatctgacca aaaactaggt accatgggtct agccacattt 60
nnnnnnnnnn nnnnnnnnnn atagtcataa gatttggttt tctcttactg aaatgcggta 120
taattgggat ttattcaaat cttcttttat atccatccca taaagttttg taatttttcc 180
ttattggcct aacaattcct gtaaaagatta tttgtaagta gttaataatt tttgtaata 240
tgatgagttt aacactttta ttttattaca tgtgctcata tgttggtgct gacactaaga 300
aacctatagg tcaaaaagatg caaactaggg ccacatgagc acggcagcac ccagccagga 360
ctctgctgca gctgcccgtt gtagatggag ctctctgtct ccaaagaagc acaggcctgt 420
tggtcttctg gttgtgctac agtaaaatga acctgggttt tctgaacatg tgggtgaatg 480
tcagtttgcta cctancgttt cacttgatgat attgatttta ttaactctta agttatgtga 540
atgttaacta tttctcatat ctatatctta ttcaaaatct ggg 583

```

<210> 8  
 <211> 118  
 <212> DNA  
 <213> Homo sapiens

```

<400> 8
gaaaaaatc ctgaaatatt caaatgttta ccttactaaa ggaattctta atctgctgat 60

```

tatgtttctt aattgtacca tgaattacaa aaacctattg gcaagatcag tcttattt 118

<210> 9  
 <211> 502  
 <212> DNA  
 <213> Homo sapiens

<400> 9  
 tccacttttt aaaaagagtt agttatatgt agcatagtaa aaatatttct tttgaaatat 60  
 actattatat ccatttattt aagttctaag agactatact gcctaaaaaa atagagttga 120  
 tcttagtcaa atcacttaaa tcaggaacct ttaagattca gtttaatcag ttttgatat 180  
 aaagtattag acttactttc accatttttg cccaaaaaca aacttctgta cttcattaaa 240  
 taacatcaac aaaagtaaaa agctaaccac aacctaggaa aaccattttc accgtaacag 300  
 ttatgtatat taactatatt atatactttt aaattagtaa aaaaaagggt aaaagacatg 360  
 aacaattatc aaaggaaaaa aatcctgaaa tattcaaatg tttaccttac taaaggaatt 420  
 cttaatctgc tgattatgtt tcttaattgt accatgaatt acaaaaacct attggcaaga 480  
 tcagtcttat ttaaaaaaaa aa 502

<210> 10  
 <211> 390  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (83)..(211)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (214)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (300)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (304)..(324)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (368)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (381)  
 <223> a, c, g or t

<400> 10  
 cacatatatc cagtaacagg gctgtgcgtg taccctctaa atctacaaaa ataaaaagtc 60  
 ctcaagtttg acttattgtt tannnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120  
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180  
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nacnctatt cctttctgga gactctaggg 240  
 gaatctcttt tcttgccctt tctggcttct agaacctgcc tacattcttt ggctagtggg 300

```

cccnnnnnnnn nnnnnnnnnn nnnngtggct ggtcagtcct tctctgatgt tatctttctg 360
gttctgancc ttccatctcc ntcttcgcca 390

```

```

<210> 11
<211> 266
<212> DNA
<213> Homo sapiens

```

```

<400> 11
cttcagtggg aagagttctc atattaaatt gaattggtag ttaagaaagg gctttaacta 60
ggcaaacct tgggagtttt gaggtccctg tggacttgcc tggcatgcct gggagatttg 120
taggtatctt ctttctacac ggatttgtgt tctgccctca aggaaaatag ttcactttga 180
ccactgtaaa tgatgtagta tttaaacaaa aggaaaagca cttcattgtc tgctctaaaa 240
ctaaaatggt aagaaagaga ggtggc 266

```

```

<210> 12
<211> 380
<212> DNA
<213> Homo sapiens

```

```

<400> 12
tgtcatggga agtggcggcc actgccacat ctgctatgag gctcttcgcc atcgttggct 60
gctggaaatt tggatacagt aaatgggtata taaggcttct ttttgcttgt gcaccagaag 120
tctttgtccc agcctccaga tcagcagtca gtactccct ctccaacct gtaggaaagta 180
cttgtgaaaa gttatctatc cctggtctga gtgggagggt cttaacctca ttgatgtttt 240
agtgtgactt gtctacattt gtgtgctccc ttcgtcatct gcagaggata tgagaaaaga 300
aacaatatga caaaaagtgg agatagcgcc ttcctattca ttcttcattt ggtatgggta 360
tagttaagag aggtgagcca 380

```

```

<210> 13
<211> 871
<212> DNA
<213> Homo sapiens

```

```

<400> 13
tttttgctaa gacaagatct actttgtcac ccaagctgga gtgtagtgge gcgatcatgg 60
ctcattgcag cctcggtctc ccaggctcaa gcaatccttc cacctcagcc tccagactag 120
ctggaactac aggcgcatac caccaagcct ggctaatttt ttaattaaga cagggtctac 180
tatgtggccc aggctggctc tgaactcctg gactcaagca attctccac tctggcctcc 240
caaagtgata ggattacagg catgagccac cagcggcagg ctggctcacc tctcttaact 300
ataaccatac caaatgaaga atgaatagga aggcgctatc tccacttttt gttcatttgt 360
ttcttttctc atatcctctg cagatgacga agggagcaca caaatgtaga caagtcacac 420
taaaacatca atgagggtta gaacctccca ctgagaccag gtagatagata ctttcacagt 480
acttctacag ttggagaggg agtactactc tgatctggag gctggacaaa acttctgtgc 540
acaagcaaaa aaagcctatt accattacct gtatccaaat tccagcagca acgatggcga 600
agagcctcat agcagatgtg gcagtggcgg ccacttccca tgacaaaagg aggtaggcat 660
gattcatctc tagtgcattg ggacgggtgc cttttgcccc aaatcaactc aagacccctt 720
tccaccatat gtctgtatgt aactctaaat gcattcttaa gacttaagaa taaaaagcca 780
ctgctacctg gcaagtggat gaggcaaaag agaagacata ccccaaagaa ctatagcact 840
ctgctccaaa ttacagaact ttctaaacgt c 871

```

```

<210> 14
<211> 411
<212> DNA
<213> Homo sapiens

```

```

<400> 14

```

```

gaacaaacct ggaatatgt gaactattga tgacaatgct ttgagggcat ctttggaac 60
caaaatgtaa atataactaa ttattttttc acttattgtt agcatttagt attttatgca 120
taaaaaacttt ttaccaaat aaattttgga agtttaaatt ccacaaatga tactaatgaa 180
agtataaatc attttgggtt gttttttaa aaattatgtt tcaatctgtc attattggaa 240
taaagtgtat aaactgcatg ttataaaacg gctttacaca tatataactc atgaactcaa 300
gagaaataat attttttagga aagaagagta tctctagatt ttttaataata attaaatttc 360
tttaaaagac tgaagataga aaagagaaat taaactatgt attgacttaa t 411

```

<210> 15  
 <211> 737  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (657)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (709)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (716)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (721)  
 <223> a, c, g or t

```

<400> 15
atcaaaacaag gcaaggagaa ctcaagaaag tctggtcact tcagttagtg ttggtgtgtg 60
aggagaatca agtgagaagt gcttttagaa acatccatgc aggtatgaag gagctctaac 120
atgccagggtg agtggacaag gcttggggaa ggttgacgat acctagagac atgtcagcct 180
ttgggtcaga gctgcctaa tctcatgggg aactgatggg tggaaataggc actctaggct 240
cccctggca tgggcagggt tgcagagaat tcagaaattt gtgtgtgttc acggagggga 300
ggccaagctt ctgactgttg aagttagggt cagaatggag aggccagcat gtgtgtgagg 360
gactctgaga tgggaggctg acaattagga caatgggctt tgcccatgga gccagaatgt 420
taccctgcga ggggtgaatgg cactgggcca catatctcag tggtcagatc caggatctca 480
gcagtcactc ctatcaatca gagatcggac catggcaggc aactcagca aagtccccca 540
tgatccagaa gacatgtgcg aattctgcat catcttcccc tccataatcc tgaggacagt 600
gagagccaag gtaaggacct tgacacacag attcgttacc aggaggaatt ctttgcntac 660
tgaatcattc tgaatatatc ggatttgtct aaatagatcc cacctgcctt cccatntact 720
nttctccagt aaagagg
737

```

<210> 16  
 <211> 1082  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (90)  
 <223> a, c, g or t

<220>

<221> unsure  
 <222> (1002)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (1054)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (1061)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (1066)  
 <223> a, c, g or t

<400> 16  
 ttatgggtag atattatgat gacgtatata ttctacagat ggagaaacta tgacctagag 60  
 tgaccccaat gtcaccagct aagggtgggn ctgggtttaa accaggccag tctagcttgt 120  
 aacctctatg tcatattgga atagactcta gcacagtggg tatagggtcc tactaaaaga 180  
 gtgttaaaac catgacagct atgctttaag aatacatact tgaatgtaag tggttagccct 240  
 gacatccaat ttgtcttcct aacgcctagt atataatatg gctcaagaaa tgtttattaa 300  
 atatgtgttt catttttttt ttcttttttg agatggagta tcaactctgtc acccaggctg 360  
 gagaactcaa gaaagtctgg tcaacttcagt gagtgttggg gtgtgaggag aatcaagtga 420  
 gaagtgtctt tagaaacatc catgcaggta tgaaggagct ctaacatgcc aggtgagtgg 480  
 acaaggcttg gggaggttg acgataccta gagacatgtc agcctttggg tcagagcctg 540  
 cctaatactca tggggaactg atgggtggaa taggcactct aggcctcccc tggcatgggc 600  
 aggggtgcag agaattcaga aatttgtgtg tgttcacgga ggggaggcca agcttctgac 660  
 tggtgaagtt agggtcagaa tggagaggcc agcatgtgtg tgagggactc tgagatggga 720  
 ggctgacaat taggacaatg ggctttgccc atggagccag aatgttaccg tgcgaggggtg 780  
 aatggcactg ggccacatat ctcaagtgtc agatccagga tctcagcagt cactcctatc 840  
 aatcagagat cggaccatgg caggcacact cagcaaagtc ccccatgatc cagaagacat 900  
 gtgcgaattc tgcacatctt tcccctccat aatcctgagg acagtgagag ccaaggtaag 960  
 gaccttgaca cacagattcg ttaccaggag gaattctttg cntactgaat cattctgaat 1020  
 atatcggtt tgtctaaata gateccacct gccntcccat ntactnttct ccagtaaaga 1080  
 gg 1082

<210> 17  
 <211> 128  
 <212> DNA  
 <213> Homo sapiens

<400> 17  
 gtttttgcaa ataagagcac taaaagact aaaccattcc tcggtgcctg gaagaggctg 60  
 tttgcatttt agttaccctg ctgttcataa catctctaag aaaatgtagg ggccaccctg 120  
 ggcgcagt 128

<210> 18  
 <211> 465  
 <212> DNA  
 <213> Homo sapiens

<400> 18  
 gaaggagaga gtgagtgagc aaatgactga aagcagccag cattgccggg accctgccgc 60  
 atctcactact ggggtgctgca caggcagcgt tcctcctctt gtggcgcttg ggaaatagac 120  
 attaatcaca cacaacaaaa aacgatggca aattgtaatg agggctatga aagggagtga 180



```

agggcaagat cactgagggg tgggactcag aagaatcctt tctgaggaca tgacctgtca 240
gctaagacgg aaggaggact tcagttgaga gaacaagtca attaatactc cgtggcctct 300
tccactcact ctcatagaatt tgccaagtcg cctggggaga atggggcctg aactctggac 360
cctgaggtct ttctccatgc actgaaagca gccccccag ccctgggccc ctccctgcga 420
cttggtgact tgactgccct attctagcat catggaaaac acatg 465

```

&lt;210&gt; 19

&lt;211&gt; 539

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 19

```

gtaaaaagga aatttgttgt cattctgtgg aatagagtga ttaagcagct tgcctaaact 60
ccccagaca gggatgcagt tggagctaga aagttaaccg aggcctggct gacaaggcct 120
cattctccat cagctgggtca aagggaacagg agctctgcat cctgtcctgc tcagtagagg 180
aggaaaggag acagctcctc cgggaaatct agaggaattt gcattctctg cctgagctgg 240
cttggggcca gctttgtgga atcctgggtgc cagttctttt tctgcttcag gggctatttt 300
ttaaaaaaat cagtagttac attttatgtt taccaataga tttatatagc aaatgatatt 360
tgttttttat ttaaagccac aatatcaagt gtctttttta aaatataaat aataatcctc 420
atggtatgca gatgtagcag aaaattgtgc aggtggtatg tgggcaactg gatttgggga 480
aatgctgctg catgtcatgc actctccata ggtaggtttt cccttttatt tcctccctc 539

```

&lt;210&gt; 20

&lt;211&gt; 641

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 20

```

gtaaaaagga aatttgttgt cattctgtgg aatagagtga ttaagcagct tgcctaaact 60
ccccagaca gggatgcagt tggagctaga aagttaaccg aggcctggct gacaaggcct 120
cattctccat cagctgggtca aagggaacagg agctctgcat cctgtcctgc tcagtagagg 180
aggaaaggag acagctcctc cgggaaatct agaggaattt gcattctctg cctgagctgg 240
cttggggcca gctttgtgga atcctgggtgc cagttctttt tctgcttcag gggctatttt 300
ttaaaaaaat cagtagttac attttatgtt taccaataga tttatatagc aaatgatatt 360
tgttttttat ttaaagccac aatatcaagt gtctttttta aaatataaat aataatcctc 420
atggtatgca gatgtagcag aaaattgtgc aggtggtatg tgggcaactg gatttgggga 480
aatgctgctg catgtcatgc acctctccat aggtaggttt tccccctta ttttctccct 540
cttttcagca aaactctccc ctctacttcc ccactgaaaa atagcatggt gagttaaact 600
agttagtctg atcatggtgt gggcatggtg gttgggttag a 641

```

&lt;210&gt; 21

&lt;211&gt; 406

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (379)

&lt;223&gt; a, c, g or t

&lt;400&gt; 21

```

ctggtagtca ggggcaaact caggccagaa cccaagtctc ctgactacta ttccagtgtc 60
ctttccaccg tgctctggaa gaatacggca caagcataag agcgtagtag tcatccctgt 120
acattcatgc agccagcctt acctagagtc acagtcaatt gtggccaaact tggcaagatt 180
tgaacatcac tgataagcaa tctttctctc aatgctgcat ctctccagct tggtctttcc 240
ctaccatccc ccacgtatga cttaaagttat agcattgact gaaatctttg ggattaaagc 300
cctgtgatct gactgagaaa aacctgttga gccattacct acaattttaca caaacaatt 360
tcttcgattt gtcttttang gctggccgga aggcatttac atttga 406

```

<210> 22  
 <211> 467  
 <212> DNA  
 <213> Homo sapiens

<400> 22  
 cagacagaac agatagatag aagaaaagaa aagggtcact tggcactagg tcttcacagg 60  
 taaagattca gagtgtgata ggaagcacag gtcaggcac ccgggtctaa tcaatgacaa 120  
 tctcgcttct aggccttttg gtggcatttt ctagtctacc tctaagctct agggaatcgt 180  
 gtggctaaaa tcttcctcc tgctgagact cagagaatac catgttggcc aagatctcta 240  
 aaacaatcaa acctggcagt attgagttac ctctctcta tcataaagtc tttcctcact 300  
 tcctccttat tgtgaacttt cttaagaagt gagtccagga ggaagcagt acatgaattt 360  
 attaacttga ctgagacttc taaagacaac acaaactggg cgccccattc agagagtgac 420  
 agggaaaccc cgtggcataa ttagttacta cgagtttcca aatagga 467

<210> 23  
 <211> 1328  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (300)..(472)  
 <223> a, c, g or t

<400> 23  
 gcacagattt agccttggtt tttttttctg ggaagtataa aagacttttg tgttctgtct 60  
 ttttggtttc aatttctctc tagaggaatt taaaaccgga tatctccatc ttaaagttct 120  
 tgagcaagtc tgtcaagggtg tccatatttc ttaccctgtt cctctcagca tcgaagtgtc 180  
 atctctgtta cactcatgtt tgctgttcac aatggagtac taatgaaata gcaaaattaa 240  
 gctaccggca tgggtgcta aactgaaact aaaaatcggt ttggagcttt tctggttggg 300  
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360  
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420  
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnaagaaaag 480  
 aaaaggggtc cttggcacta ggtcttcaca ggtaaagatt cagagtgtga taggaagcac 540  
 aggctcaggc acccggtctc aatcaatgac aatctcgctt ctaggccttt tgggtggcatt 600  
 ttctagtcta cctctaagct ctagggaatc gtgtggctaa aatcttcct cctgctgaga 660  
 ctgagagaat accatgttgg ccaagatctc taaaacaatc aaacctggca gtattgagtt 720  
 accttctctc tatcataaag tctttcctca ctctcctct attgtgaact ttcttaagaa 780  
 gtgagtccag gaggaagcag tgacatgaat ttattaactt gactcagact tctaaagaca 840  
 acacaaactg ggcgccccat tcagagagtg acagggaaac cccgtggcat aattagttac 900  
 ctacgagttt ccaaatagga tttggaagga gacatacaac taggtcgccg gcgtggcaca 960  
 tggcttccct gaagccagca ttgcttggcc aaggaagctt tgcagaacag atgagatttc 1020  
 agctgggact tgcagccaag tgggatttgg ccttttgggg agaagggaaa gggcattcaa 1080  
 aggccaggga cagagtatgg tcaaaggcat ggagatgagg aagaggggac cagagcagag 1140  
 ggtcaggttg gaaagcgagt tggggtcaat ctgcaaaggg gctgacgtgc caggtaaaaa 1200  
 acaggagcac cgtttagttt tgtcggatca tttcagggtg aagggcagtg ggaatgttgg 1260  
 agaaaacact ttttgggtgc gttacattga atctgctcat ctataagaat aaaactttat 1320  
 ttcataga 1328

<210> 24  
 <211> 550  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure

&lt;222&gt; (352)..(371)

&lt;223&gt; a, c, g or t

&lt;400&gt; 24

```

ctcatacctc ccctgggtcca gatcatgatt caggctctttg ttctaggatt cctggcttat 60
tatctgggtt atttcaataa caaggacaat aagtcattgg tatgattttt ctgtttcatg 120
gagtgaagtga acattttatc tcattccagg aatttggttt ttccaacta ttgttgcttt 180
ttgggttggg tttaaataatt cctttacca gaaattcatt cattagtcta cttttcagc 240
tttattagca taggagtcct aataacattt tgtgtatggt ttcatcacct agtgattggt 300
tgctaattcc tcaccttttc ttttaggtcac tgttctttat acattgggta tnnnnnnnnn 360
nnnnnnnnnn ngcatttaag tttttacttt ttatgagaca aatgtatttg cgttccatag 420
atgtcagtta gaaatgtttt caacatcatg gttctctaca aactttgtga tttcagttac 480
atttccacat tgactcaaca gttatttaat agtgagctct ttttttttta agacgtagtc 540
tgactctggt                                     550

```

&lt;210&gt; 25

&lt;211&gt; 150

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 25

```

gattcagccc gtgatccttg actggatcct ggggtgaaagc aaaagcagct ctaaaggaca 60
ctttgcagac taaatgttag ctaacaccat tgtatcagtg agaaagtgca gagtgtgggtg 120
agtccattga ggctctgtag aagaaagtc                                     150

```

&lt;210&gt; 26

&lt;211&gt; 192

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 26

```

aaaaagtcaa tgtcataaaa gacaaagaaa ggctgaagaa gtgattcagc ccgtgatcct 60
tgactggatc ctgggtgaaa gcaaaagcag ctctaaagga cactttgcag actaaatggt 120
agctaaccac attgtatcag tgagaaagtg cagagtgtgg tgagtccatt gaggctctgt 180
agaagaaagt cc                                     192

```

&lt;210&gt; 27

&lt;211&gt; 747

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 27

```

gagctttgca gggatttagc ttttctcagg gccacctgcc ctgaggcttc ctgggcccctc 60
atacttcttc ttgtttatat cttatctgcc tttgggggaa tgaccttaga ggaattgggtg 120
tgagtaagcc atgaggttct tgggccacct ccatccagcc aagggcagct ggcagctggg 180
cacttacatc cagcaaggca gaagcaaccc tggctttgaa gtcagactgc tagggtgagt 240
ctgaatggcc tcggggaaag ttccctctga gccttcgttt ttttcaactg tgaaggcgat 300
agtctcgct agcttgaggg tttatcaggg ggattcagtg agaacctcat ttgaagcagc 360
tgcttttagtt cctaacacct aataaatggt taaccactta ccctcctctc ccaccaccct 420
ttcaactttg aacctcttcc tccatgtcat cccttcttaa ggcgctgacc ttttggccac 480
aaagaatggc tctttttggt cccatcagga ctagaattct tatctttttg ttgcttggcc 540
ctggtaatca aagaaccacc aacacatttg caaggcatct ccagccttct cgttctggcc 600
gccctctct gtcttaggga gagtgtctata ctggcatggg gatgagatga acgaaagggc 660
agtctctggc tgttttctgc tgatgaggat gtgctgagca gcctcctgca aatgagaagc 720
agggaaaaga ccaaactagc ttagctc                                     747

```

&lt;210&gt; 28

<211> 184  
 <212> DNA  
 <213> Homo sapiens

<400> 28  
 taagctcgga attcggtcgc aggcattgcca gtctttgggg catatggatg gtggatgtgg 60  
 cttgcctttt caccctctga cattgttgat gagcagaaac ttccaatttt gatgtagtcc 120  
 acaatattaa tatttctcat aggccaccac acccagccta tggtatcttt tagaagcttt 180  
 attg 184

<210> 29  
 <211> 217  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (97)..(161)  
 <223> a, c, g or t

<400> 29  
 ctcaattatt ctaggaatct atgctgaata tgcctctaac aatacaaat atgtattact 60  
 taatgttatt aattatagta ttatttaatt tgcgaannnn nnnnnnnnnn nnnnnnnnnn 120  
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ntagccatat aatggagtac 180  
 catgcagata aatcagaaac tagagattgg ttacctg 217

<210> 30  
 <211> 543  
 <212> DNA  
 <213> Homo sapiens

<400> 30  
 tgaggaagcc tgaatgggaa tgaaaaacaa ataagaatca aaagcagagg gctgtgtctt 60  
 ttatttttgt attcttacac cgtatgaact ttattaagaa ctaaaatcat gtgatggtaa 120  
 cccatgggca cagctgagt acaaagttgg aaaattgtac cagcatcatt aactgggggtt 180  
 gctttgtgtt aacatttgtt atcattcaca gatgtcaaca ataagagaac acatctctct 240  
 atatataatt gttacaaaca ttttaaatta taaggaaaag aaaaaaaaag atgcaaaagt 300  
 tcaaagacta aactcacaac atcctaccga cagagaatac ttaggatagc taagatctca 360  
 gtgttcattc aaacttctaa actcaaagac tcggatttat gcaggaagtt gaacatgtct 420  
 tgtgtcacca gatctgtcag tgaagctgat tatagtcttg ggataaattt gagagtaatg 480  
 tgacagtggg ctgggaactc tgactgtggg ctctgccctc tggatgcaga gactccaagg 540  
 cac 543

<210> 31  
 <211> 283  
 <212> DNA  
 <213> Homo sapiens

<400> 31  
 gcttcaagct tctgtctctc ctcttttgcc atattattga gcctggaatc tgagtgggag 60  
 aggacgacag agggctctggg cacaaggaag ccattgattg aggccattac tgcaatcaac 120  
 ccaccaccaa taaaaagcac tggaggagag ggcttgacat agacacaaaa cataaaggaa 180  
 gggggtgaaa ggaaggaaag agattgagga aaaaaataa aaataaaaaga tggctgggta 240  
 agggaagaag agatagggaa gagagacaga aaaggtagaa tgc 283

<210> 32  
 <211> 418

<212> DNA  
<213> Homo sapiens

<400> 32  
aaaaaatcct tcatgtgtct tgtatacatc tgcagaagac cagtagtgta cattttcttg 60  
gctgtagcaa gggccaaga aaggaagtag taagcagttt ttcaagtctc tctctctctt 120  
tttttatttt gttggcttca atgcttcctg ctctctctct ttgccatatt attgagcctg 180  
gaatctgagt gggagaggac gacagagggt ctgggcacaa ggaagccatt gattgaggcc 240  
attactgcaa tcaaccacc accaataaaa agcactggag gagagggctt gacatagaca 300  
caaaacataa aggaaggggg tgaaaggaag gaaagagatt gaggaaaaaa aataaaaaata 360  
aaagatggct gggtaaggga agaagagata gggaagagag acagaaaagg tagaatgc 418

<210> 33  
<211> 172  
<212> DNA  
<213> Homo sapiens

<400> 33  
cagactggga ctctggaaaa tcctaaagca ttatagaact tggggcttgt cctttgactt 60  
catgggtttc aaaccacgca tggtgaccca gtagtgggct gtccaatcaa ctgagctctt 120  
gaaattggaa tagaataaaa tagaatatg agcatattcc catctataga aa 172

<210> 34  
<211> 128  
<212> DNA  
<213> Homo sapiens

<400> 34  
ggcctccgat tgtccacag ttagttgttc ctggaggca cccctcctgc tgctccttgg 60  
atactccagg ggccgaggag ccgagactca ctggagtgtg ggcatggcca tccagagagc 120  
tctgatca 128

<210> 35  
<211> 619  
<212> DNA  
<213> Homo sapiens

<220>  
<221> unsure  
<222> (128)..(416)  
<223> a, c, g or t

<400> 35  
ggcctccgat tgtccacag ttagttgttc ctggaggca cccctcctgc tgctccttgg 60  
atactccagg gccgaggagc cgagactcac tggagtgtgg gcatggccat ccagagagct 120  
ctgatcannnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnngctt 420  
tggtccttgg ggggtgctga aaaagcaaaa ccaggtctgt ggggtagaag gcgccctggc 480  
cacacacagg cattgccgcc tctgggggtcc gcagagtctg tgtgacaacc tggtcactcc 540  
gatctagcag cgtatttgaa tgaatgagtg acagcttaat gaagtagcca agtaccttga 600  
tttgaacgta ggagccggg 619

<210> 36  
<211> 356

<212> DNA  
<213> Homo sapiens

<400> 36  
cgacagataa gtcagatatac gaatatagac attaaaagat ctggggcact aggcgtgtacc 60  
ctgtttattgt cagtggctct ttagtcctta aacaagggtc ttgcctccta cttttttttt 120  
gttatggttag aaataaatgc ccaccgagt tttcatcact cactattatt ctatcttttg 180  
tgtgcctgta ccatgttccc ttaacaatcc tcaattatga aacatttagg cagtttataa 240  
acaatactgc aatgaacaac ctagtgcata ctttttttgt gtgttcttct tttattattt 300  
cctagaagtg agccctagaa atggagttcc tgagtcaaaa tgacacattt tatagc 356

<210> 37  
<211> 158  
<212> DNA  
<213> Homo sapiens

<400> 37  
aaggaattag attccacatc tcaatctaag gagcagcaca aatatgcaga gaggaaagga 60  
attgattgtg gccctctttg aaaactatct caggccatcc ttggggcact tcaattcata 120  
gtcttctcct atgcaaaaata cactcacctc ttgcattt 158

<210> 38  
<211> 585  
<212> DNA  
<213> Homo sapiens

<220>  
<221> unsure  
<222> (159)..(279)  
<223> a, c, g or t

<400> 38  
aaggaattag attccacatc tcaatctaag gagcagcaca aatatgcaga gaggaaagga 60  
attgattgtg gccctctttg aaaactatct caggccatcc ttggggcact tcaattcata 120  
gtcttctcct atgcaaaaata cactcacctc ttgcattt nnnnnnnnnn nnnnnnnnnn 180  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300  
gtgtcagatg agctatatcc tcatctggag gcttgacagg ggaagagcct acttccaagt 360  
tcattaggtt ggtggttagaa ttactccat tgtggatgta tgactgagga ccctggcttt 420  
ttgctagcca tcagaagagg ccagtcttgg gtcctagaga ccacctgtgg ctcccttaca 480  
atgtgggctt tctcaacatg gctacttact gcatgaagcc agcaaaaaga atctcccagt 540  
ccagtatgct aagacagagc cttgttataa cataagtcca ccctc 585

<210> 39  
<211> 295  
<212> DNA  
<213> Homo sapiens

<400> 39  
ggggggggca gtagtttctg aagagagagc taaactgcat gagcagatgc ttagccaatt 60  
tctaaaaatg gaatgggagg tagaaatttc acagggtggtt gctggtttgc agcatttcca 120  
cactactagga tacatcatca caagatgttg tctgccagct ggtgctataa ctgctagtaa 180  
agccacttgc ttctgaatgc atggtgatag tagtgaatcc ctaatgtcag tgcaatgctt 240  
tacttatttg ctataaaatc ctttcatagt cagaagcact gttgtgttcc tggca 295

<210> 40  
<211> 302

<212> DNA  
<213> Homo sapiens

<400> 40  
gcgggggggg gggcagtagt ttctgaagag agagctaaac tgcattgagca gatgcttagc 60  
caattttctaa aaatggaatg ggaggtagaa atttcacagg tgggttgctgg tttgcagcat 120  
ttccacatac taggatacat catcacaaga tgttgctctgc cagctgggtgc tataactgct 180  
agtaaagcca cttgcttctg aatgcatggt gatagtagtg aatcccttaa tgtcagtgc 240  
atgcttttact tatttgctat aaaatctctt tcatagtcag aagcactggt gtgttcctgg 300  
ca 302

<210> 41  
<211> 346  
<212> DNA  
<213> Homo sapiens

<400> 41  
aagtaattaa cttgatcaaa ctcatTTTTac agatgaggaa actgattcca cctccatgc 60  
tcttcacctg cattctaaac tcttccaggg ccctccttac caggcagagg caaattgagg 120  
aagtggacac agcatttcct ttcttggttg tttgacatgc aaagcacttt agactatatt 180  
tagtacctaa ttgatgtggc agcagggggc gcctgggatg ttgtggcatc atttttgctc 240  
tcaatgagac acgataggga tgggttggtg gtgggttcaa aactaaagac cctccagcag 300  
agcctgtcaa gtaaaacaag ggtgactgct tgggtgccat accagg 346

<210> 42  
<211> 468  
<212> DNA  
<213> Homo sapiens

<400> 42  
aagtaattaa cttgatcaaa ctcatTTTTac agatgaggaa actgattcca cctccatgc 60  
tcttcacctg cattctaaac tcttccaggg ccctccttac caggcagagg caaattgagg 120  
aagtggacac agcatttcct ttcttggttg tttgacatgc aaagcacttt agactatatt 180  
tagtacctaa ttgatgtggc agcagggggc gcctgggatg ttgtggcatc atttttgctc 240  
tcaatgagac acgataggga tgggttggtg gtgggttcaa aactaaagac cctccagcag 300  
agcctgtcaa gtaaaacaag gttgactgct tgggtgccat accaggcaca gggttagcatg 360  
aaacaaagtg tagtgtccaa ggagagggag cagggtgtct cctttgggtg agctttgcaa 420  
ggggacttggt gacttggtg gaaaagggtt tttttttagt tgtatgtt 468

<210> 43  
<211> 107  
<212> DNA  
<213> Homo sapiens

<400> 43  
ttcacctgct tgtgtgaatt gtggctttta atgtattcct gtcaattcca tatattttta 60  
aaatgttgct tttagagtat gtgcaagttt ggggcatttt tgagggc 107

<210> 44  
<211> 352  
<212> DNA  
<213> Homo sapiens

<400> 44  
gaacatgatt gagttagaaa ccagtgtggc ctgggactgg gaagctcatt aaaggaattg 60  
ggacttaaac tgggaagggc aagttggctc tagatccata gaaactgaag acaggggaag 120  
agagagatgg tattatagat ggaagaaggg gcagtgggtc atggaataaa tattgggtgag 180

```

caggggagca aaccaaaggg gtaattggga gattctgagt tttcaaggct attaaaatgc 240
agttccaggc cctagggagg agagttccag actgttttct ctacactgct ataattcctt 300
acactgctgg gagcagtttc tttgacatac tttgcaactg cagagggctt tt 352

```

```

<210> 45
<211> 356
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (98)
<223> a, c, g or t

```

```

<400> 45
gaacatgatt gagttagaaa ccagtgtggc ctgggactgg gaagctcatt aaaggaattg 60
ggacttaaaa ctgggaaggg caagttagct cctagatncc catagaaact gaagacaggg 120
gaagagagag atggtattat agatggaaga aggggcagtg ggtcatggaa taaatattgg 180
tgagcagggg agcaaaccac aggggtaatt gggagattct gagttttcaa ggctattaaa 240
atgcagttcc aggccctagg gaggagagtt ccagactggt ttctctacac tgctataatt 300
ccttacactg ctggggagcag tttctttgac atactttgca actgcagagg gctttt 356

```

```

<210> 46
<211> 482
<212> DNA
<213> Homo sapiens

```

```

<400> 46
ttgttgaaat tttgtttgac tgcttttagta caggagtata ttccccaaga caagagacct 60
gagagctttt ccctgggttaa gataccaagg atgatttcca aatttttagac atccttcccc 120
ttgttccacc aatttttttt ttcttctggg aaaatagcca ggatgattgc aaacataag 180
cttgtaaaaa ggcaaaactc catggatgta agaaagtaaa tttcttgagg gccacacca 240
tgataacgct ggaattttca ttttaattcct aactcatttt ttgttgtttt tgttttttta 300
aactcaaatg tgtctcttta attgaggtca cttacttggt tgggagatta atattctggt 360
ggggaaaactt tctttttaga gtttatattg ttttattcct tcagtcactc agtattacta 420
atggggtagc ttttgaatt ttccatcccc cccactttca gattactttt ggtctttttt 480
tt 482

```

```

<210> 47
<211> 462
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (380)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (423)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (451)
<223> a, c, g or t

```



&lt;400&gt; 47

```

gggaggccct gcttcctgcg agctgtcccg gcaggacaga gactettccc gccgcggccc 60
tgccattcca ggctgaggct gtgagcagca ccatgacaag ctccagccgc agtggctctc 120
aacagtgtgg gtctctgacc acccgacgag ctggaagtgc agaccgctga cctcccttga 180
gaacctactg ggttcttgca gtaggtcctc cagcgggtgc taaacacgcc actcagatga 240
ttctatgcac catcacattg gaaacttttt tcattgactg ttacttaatg agaagacttc 300
cctccgggat ggttctgaag ctcccttcat aggagcaage ctttggcggg agagcactga 360
gcagacgtgc agcatctttn ctggcttcta ccgaaacacc atggatccag acgtgggttt 420
tgngtctgca cgtggaagcc agccctgcgt ngggtgagcc tg 462

```

&lt;210&gt; 48

&lt;211&gt; 1609

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 48

```

atgaggctgc ggtcatcagg gctggaggag ttagaggagg gcaggaagga ccttgccatt 60
ctcttatcaa ggaactgcc a ggacctctgt gaagctgagg ccttcagcga gctcctccca 120
ggaccgtcgg agagtcccag gaatctggct gtgctgattg gtacagtctt cttcagattt 180
attctatata aagtaagcat attgtcaacc ttccctcgctc ctttcaagca cctgagtcc 240
ggcatcacia acacggagga tgacgacacc ctacgtacca gcagcgcgga ggtgaaggag 300
aaccgcaacg tgggcaacct ggccgcgcgg ccaccgccct ccggggaccg ggcccggggc 360
ggcgcgcccg gcgcgaagag gaagcggccg ctggaggagg ggaatggggg ccacttgtgc 420
aaactgcagc tgggtctgga gaagctgtcg tggctcggtg cgcccaagaa cgcgctggtg 480
cagctgcacg agctgaggc gggcctgcag taccggacag tgtcgagac gggcccgggtg 540
catgccccgg tcttcgcggt agcgggtggag gtgaacgggc tcacgttcga gggcacaggc 600
cccaccaaga agaaggccaa gatgcgcgcg gcggagctgg cactcaggtc cttcgtgcag 660
ttccccaaac cctgccaggc gcacctggcc atgggcgggg gcccgggccc cggcacaggc 720
ttcacctcgg accaggccga tttccccgac acgctcttcc aggagttcga gccccggcg 780
ccgcgccccg gactcgcggg agggcgcccc ggggacgccc cgcttctgtc cgcggcctac 840
ggcgacagcg ggctgctgtg ccgcgcgctg gacctggtgg gcccgacccc cgccaccccc 900
gcggcccccg gcgagcgcaa ccccggtgtg ctgctgaacc gcctgcgcgc cgggctgcgc 960
tacgtgtgtc tggcagaacc ggccgagcgg cgcgcgcgga gcttcgtgat ggccgtgagc 1020
gtggacggca ggacgttcga gggctcgggg cgcagcaaga agctggcccc ggggtcaggcc 1080
gcgcaggccg cactgcagga gctgttcgac atccagatgc ccggccacgc gcccggcagg 1140
gccaggagga gcgcaatgcc gcagggtgtg cttccacgtg caggaccaca aaaccacgtc 1200
tggatccatg gtgtttcggt agaagccagc aaagtgtgtg cacgtctgct cagtgtcttc 1260
ccgccaaagg cttgctccta tgaaggaagc ttcagaacca tcccggaggg aagtcttctc 1320
attaagtaac agtcaatgaa aaaagtttcc aatgtgatgg tgcatagaat catctgagt 1380
gcgtgtttag acaccgctga ggagcctact gcaagaaccc agtaggttct caagggagg 1440
cagcggctcg cacttccagc tcgtcgggtg gtcagagacc cacactgttg agagccactg 1500
cggctggagc ttgtcatggt gctgctcaca gcctcagcct ggaatggcag ggccgcggcg 1560
ggaagagtct ctgtcctgcc gggacagctc gcaggaagca gggcctccc 1609

```

&lt;210&gt; 49

&lt;211&gt; 272

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 49

```

gctcccacca cggttaaaat tgagctaagt gagtatcaaa cacacttggt cactttttaa 60
agagaaagca cattacctga aaatggcatc tttcctcttg tcaactcctg ccaaaagaaa 120
gcctcaccct ctaccctctg cacacccagc aattcactat ttcaggcagc cctctggaaa 180
ctaaggacat ctaaactctaa aggtcaggat tctgtgagag aaaaccagtc caacatgctc 240
atgaaatcct aactgtgcac agggttggat gg 272

```

&lt;210&gt; 50

&lt;211&gt; 405

<212> DNA  
<213> Homo sapiens

<400> 50  
ctgatcccca cctttgtcct agatgccaaa tatgcagctc tcatgggaca gccctggggt 60  
ctgtgtgcaa tctgtgtcca catctgtctc ctgctagact ctgtctcact taggagtttt 120  
agtactgcac agcacctgga acgagcttcc aaatccactt cttccctcca ccatctcata 180  
ttaattaatc ctgccagaga gggctgcaca ggccgaactg cctgagaata gcaaagaggt 240  
tgtttcaggc ttgggaacta cagagacacc tgtaatgggg aagggatgct ctgccaagt 300  
gagccgtggg cactgctggg agccacacag gactttgcat taggtcatgc aaacccaca 360  
cagaagcagg agttggcaag ggccaggcct gcagggcccc agggga 405

<210> 51  
<211> 294  
<212> DNA  
<213> Homo sapiens

<400> 51  
gggatgcata gtgagtccaa aatcacctcc ttctcatgaa tccgggagtc cctggagctc 60  
acggagggct tcctgggtgc ttgaaggggt ggaatccagt ctggggcgcc ccatcttcc 120  
gccgcctgcg gttgctgcag ccttctgtct tcaactgtga ccctgggata ctgcccgggt 180  
gctggctgga aggctggctt cccagagcag tgaccgcgtg tggcctgctt cctgagagct 240  
ctgcgtgtga cagcatttcc ttctgtctct ataaaaacac aaaaattagc cagg 294

<210> 52  
<211> 3381  
<212> DNA  
<213> Homo sapiens

<400> 52  
atggagaagt ttctgcagat cgcgcctcac tccctggcca tegtcttggg cccggcagag 60  
gcgcggcgcg gggaaaggcc aggggcagcc cggcccgcg ccccggccca gcccggccag 120  
ctcgcgccgc accacatcgg ctacgagatc ttcgccgact tcaaagccga gaacatgcag 180  
cacttctgga acaagaaggt cacggccgcg gtggccgaga ctttcttcc gggctggatc 240  
gacgagcagg tcctgctgat ccagggcaag gaggaacatc tggaggcgct gcgcgaaggc 300  
tggacggccc ggccctgcg gcgcctctcg ggcttccaca tccgctgcct ggtgatgta 360  
tcacccatca gtatgtctcc catcagtcag tctcagttta ttccactcgg ggagatctc 420  
tgcttggcca tctcagcaat gaactcggca agaaagcctg tcacccaaga agcactgatg 480  
gagcacctga ccacgtgctt cccaggtggt ccaacgccea gccaaagaaat tctgcggcac 540  
acgctgaaca cgctggtacg ggagaggaag atctacccaa ctccagatgg ctacttcatc 600  
gtgaccccac agacttattt cataactcct tccctcataa gaactaacag taaatggtac 660  
catttggacg agaggatacc tgaccggtct cagtgcacct ctccgcaacc cgggaccatc 720  
acgccctctg cctcaggctg tgtcagggaaggacattgc cccgaaacca ctgcgactct 780  
tgccactgct gcagagaaga cgtgcacagc acgcatgcac ccaccctgca aaggaggtct 840  
gccaaaggact gcaaagaccc ttactgtccc ctttctctgt gccaggtgcc acccactgaa 900  
aagagcaaaa gtactgtaaa tttttcctat aagacagaaa ctctctcaaa acctaaagat 960  
agtgaagagc agtcaaaaaa attcgggcta aagttattcc ggttaagttt taaaaaagac 1020  
aagaccaaac agctggccaa tttttctgcc cagtttctc ctgaagagtg gccctgcca 1080  
gacgaggaca cgcagctac gatccctcgg gaagtagaga tggaaatcat taggcgcatt 1140  
aaccagacc tgaccgtgga aaatgtcatg cggcacaccg cgctcatgaa gaaactggaa 1200  
gaagaaaagg cccagaggag taaagccggg tcctctgccc atcacagcg aaggagttaa 1260  
aagagtagga ctatcggaa gtcccatgga aagtctcgg ctcacagcaa gacacgggtg 1320  
tctaaaggag acccttccga cggttcacat ctggatatcc cagctgaaag agagtatgac 1380  
ttttgtgatc ctcttaccag gagatccaac aaagccaagg agagatccag gtcgatggat 1440  
aactccaaag gccctctggg tgcttctct ctggggacgc cggaagacct tgctgaaggc 1500  
tgacgccaag acgaccagac cccagcccaa tctacattg acgacagtac ttaaggcct 1560  
gcacagaccg ttagtctcca aagggtcac atttctgcca caagctataa agaggtgtgt 1620  
attccagaga tagtcagtgg cagcaaggaa ccgtccagcg cttgcagcct ttggagcca 1680  
ggaaaaccac ccgagagttt gccatcctat ggcgaactca actcttgtcc aacaaaaa 1740

```

gccacagatg actattttcca gtgcaacacc tctactatca caagtcgagc ctgtccctcc 1800
tcaaactctca cccgaagaca cctgctgaca cattgccagg ccgatgtgag aaactggaac 1860
cgtccctcgg ggacctcggc ggcacaagcc atgcctgctt cccagcgtca gcaggagtca 1920
ggaggggaacc aggaagcctc ttttgactat tacaacgtct ctgatgatga cgactctgag 1980
gaagggggcaa acaagaacac agaggaggag aaaaatagag aggacgtagg caccatgcag 2040
tggctcctcg agcgggagaa ggaaagagac ttgcagagga aatttgaaaa gaacctcacc 2100
cttcttgctc caaaagaaac cgacagcagc agcaaccaga gagccacca ttcagcccg 2160
ctcgacagca tggacagcag cagcatcaca gtggacagtg gattcaactc cccacgatgc 2220
cctgcagctc tgaaggctga agcatcctac tcagaacacc aggagaaaca ggcatttaag 2280
atgccagaac aggagttggc ttgtttctta ggtcggaaga tcctgaacta cgatgacatg 2340
agcaacctaa acatccccag tgagattggg gggctccctg gtgtgaggct ctgcaggct 2400
gagaggctga aagctccccg gttctccctc ttctgggta tctatgaaca ttacagtgc 2460
acaaacttca ccatggcggc aaacagcagc aatggccaga caaagacagg ggccccactg 2520
cagcagggtg aggaccccc caagatgatt ccagtgcagt tgattcgagg accaagtttg 2580
agaaaacaaa ctacagctgcc agccttgccc ctcaccgtcc tcatcagcga acctgttccc 2640
ctgagaatgg cttgctcatc tatctcaggg gatacccagg cagcctctgg ctccagcgac 2700
tccctgtccc cgtcactggc cttcgaggca cagttcgtct tccctgtgaa agctgccact 2760
ctgacctccc tgctgagtat gagtgatggg gtagaatcgg cgaatccttt agatctttac 2820
tgcattgttg ctgaagttat gctgattgtg actatttctg ggctgtcaat ggatttttat 2880
gtggccctga gcacagtgtc tcaggggagg tcaagggtag agaatccggg tttatttatc 2940
catgcagagc agtactgtag ttacaaggtc aaaaggaaac caccaaaca gttattctt 3000
catagaaaaa atagtccaaa tcctggaggg aaagtggggg tcgggggaag aaggagggca 3060
cctgacacag cgctggcaga ggctgagaca gaaggaaatg ctgtcacacg cagagctctc 3120
aggaagcagg ccacagcggg tcaactgctct ggggaagccag ccttccagcc agcaccgccg 3180
cagtatccca ggggttcacag tgaagacaga aggctgcagc aaccgcaggc ggaggaaga 3240
tggggcgccc cagactggat tccaccctt caagacacca ggaagccctc cgtgagctcc 3300
agggactccc ggattcatga gaaggagggtg attttggact cactatgcat cccagcttc 3360
catgccagag agctagacta a

```

<210> 53  
 <211> 245  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (199)  
 <223> a, c, g or t

```

<400> 53
gaaaacctta agatacacag gtataagatg ttattattta ctcatccatt ctgcaaatat 60
atacccttat tacctagact ctgttgagga tagaggccgt gctttttccc acgtgtggag 120
tagagcaaca tagaagagtg ccctgcagtg ggtgctgtaa tggagatgtg taccagggtac 180
aacaggaacc taagggggna aaggaacccc tgagtttatc ggggggcacc agggaaggct 240
tcaca

```

<210> 54  
 <211> 388  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (282)  
 <223> a, c, g or t

```

<400> 54
tctctatcct gacttcatga tccaccgcc ttggctccca aagtgtggg attcagggtg 60
gagccatcat gcctgggtat ttgaaaacc ttaagataca cagggtataag atgttattat 120

```

```

ttactcatcc attctgcaaa tatatacccc tattacctag actctgttga ggatagaggc 180
cgtgcttttt ccacgtgtg gagtagagca acatagaaga gtgccctgca gtgggtgctg 240
taatggagat gtgtaccagg tacaacagga acctaagggg gnaaaggaac ccctgagttt 300
atcggggggc accagggaag gcttcacaga ggacatgatg tgagttgcca tttgaagaat 360
gagaaattgt tcttctgatg aactaaac 388

```

<210> 55  
 <211> 360  
 <212> DNA  
 <213> Homo sapiens

```

<400> 55
ccaaacagaa tctttcagaa tcctgttgtt ctggcaggta ttttttaata tttccttctt 60
acaaattttc tttttacaga tgatatgcat tatagtgaca ttaacaatca atcttagaca 120
caagtgattg tttttataaa taggatctcc tcaatattag tgatcctata ttaagaaaga 180
tagtacatgt gaaccaatgg taacccaaaa gaatttgaaa agcaataatt tagtgggagc 240
tcacttgga tataactcta tgtcatcatg tatttattta agtcatattc tatgaaatat 300
cctatttgaa agcaaggaca ccctttggtt gcaaccccaa gttacttatg cagtattcgt 360

```

<210> 56  
 <211> 1203  
 <212> DNA  
 <213> Homo sapiens

```

<400> 56
atcttgctct ggctaatttt cgtattttta gtagagacag ggtttcacca tgttggccag 60
tctgatctca aactcctgat ctcaagtgat ctgcctgcct tggcctccca aagtgttggg 120
attacaggcg ccacgccagc caacacctct tttttacgaa gtactgcatg agtaacttgg 180
ggttgcaacc aaagggtgtc cttgctttca aataggatat tcatagaat atgacttaaa 240
taaatacatg atgacataga gttatattcc aagtgagctc ccaactaaatt attgcttttc 300
aaattctttt tggttacat tggttcacat gtactatctt tcttaatata ggatcactaa 360
tattgaggag atcctattta taaaaacaat cacttggtgc taagattgat tgttaatggt 420
aatataatgc atatcatctg taaaaaggaa atttgtaaga aggaaatatt aaaaaatacc 480
tgccagaaca acaggattct gaaagattct gtttggaaaa aacaaacaaa cgaacaaaaa 540
aacgttttat ggggctagggt tttataacct tttcccagtt attttctttt gctttcttca 600
ccacgttgtg gcaggccagg tttcactaac tcaggcttcc ataacaacgg tttcagcact 660
gaccgagtgg ttccatcaaa tattaacagc tgagagagtc agtgcccttc tgcaagggct 720
ggaatgtcac aaaagcccat caagagcttt gcctcggcct ttcctgggcc ttaaatcatg 780
acaggataat gaaggaatc ttaacgggac ccgttttaga gtaaataagt tttattgggg 840
ggtccaaaga aactcccag gcctccacaa acaagcctta ttgggtacta aagaaactcc 900
ccaaacctcc atgatttagc aggagacaag acaaaggatg tgaccccgag acctggaccc 960
atthaggtta agtaaattta ctgaggttcc agaggaagag cttcagggct cagatcttat 1020
ttatagattt aaaaaagtga atcacttatg tcttttagatg aatgtacact cacatgtaga 1080
catatagctt ataattgtaa taagctctgg aaaactttgt agttttgagt tggctttggg 1140
atcatttcca ggctttctcc ccatacctgg ttacagaaat aaactccctc ctttatcagt 1200
taa 1203

```

<210> 57  
 <211> 780  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (192)..(219)  
 <223> a, c, g or t  
 <220>

<221> unsure  
 <222> (442)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (452)  
 <223> a, c, g or t

<400> 57  
 gctcgacaat tttatgacta tatttaatta atatgagcac attttagagt ggagaaaaca 60  
 aagacttcat tcatccagca aatatttatt cagtacccga agtgctagaa actatatgag 120  
 ggtagtaaat aaaatagaat attcctgtcc tactatggag ggggaatgga gaggagtg 180  
 aggaatagat gnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ttcattgtac ataaagaagt 240  
 tatttttttc tgacagtaac taacaaaggt ctgggcaaga atcagagggg gaccatttta 300  
 agaggtgggtg tttctgttga gactcaaag atagaagga tccagtgtg cagaaatcca 360  
 gggcaaggaa taggatgttt gaagcctcca tagaagaaa gcattttata gtagatcaga 420  
 aagcaataac aaaaaagaaa anagaaaaaa anccatttgg caatgtctag gaacaaaaag 480  
 gacattaacg tgggtagaat gctgtgagct aaaaagagag tagattgaaa tgaagttaa 540  
 gagaaatgga gagacagacc tcatagaatt ttgccctaaa tgaaatggga agccaggga 600  
 gtatgacaca gtcccataat aaacctgctt ctggtgcaga atggattgga attatcaagg 660  
 cagttagtga ggaatccagt tagaaggtga atacagtggg tcagtgttcc aggatggaaa 720  
 tcacagtgcac ctcaactaag aaggcagcag tagaggtaga gagaagttga tagatttgtc 780

<210> 58  
 <211> 945  
 <212> DNA  
 <213> Homo sapiens

<400> 58  
 gctcgacaat tttatgacta tatttaatta atatgagcac attttagagt ggagaaaaca 60  
 aagacttcat tcatccagca aatatttatt cagtacccga agtgctagaa actatatgag 120  
 ggtagtaaat aaaatagaat attcctgtcc tactatggag ggggaatgga gaggagtg 180  
 aggaatagat gataaacaaa caaacaaaca agcaaaactag ttcattgtac ataaagaagt 240  
 tatttttttc tgacagtaac taacaaaggt ctgggcaaga atcagagggg gaccatttta 300  
 agaggtgggtg tttctgttga gactcaaag atagaagga tccagtgtg cagaaatcca 360  
 gggcaaggaa taggatgttt gaagcctcca tagaagaaa gcattttata gtagatcaga 420  
 aagcaataac aaaaaagaaa aaagaaaaaa aaaacatttg gcaatgtcta ggaacaaaa 480  
 ggacattaac gtgggtagaa tgctgtgagc taaaaagaga gtagattgaa atgaagttaa 540  
 agagaaatgg agagacagac ctcatagaat tttgccctaa atgaaatggg aagccaggga 600  
 agtatgacac agtcccataa taaacctgct tctggtgcag aatggattgg aattatcaag 660  
 gcagttagtg aggaatccag ttagaaggtg aatacagtgg ttcagtggc caggatggaa 720  
 atcacagtga cctcaactaa gaaggcagca gtagaggtag agagaagttg atagatttgt 780  
 cagttaagtc ctgaatcacc ttgattgtta cctactctct ctctttgggt cttaattttg 840  
 tcttctgtaa actgggatca attatattaa tgccaagaga tgttcagaa atgattttag 900  
 acaagacatg agtacctgac ataaggtagg atgcagtaat ctgag 945

<210> 59  
 <211> 444  
 <212> DNA  
 <213> Homo sapiens

<400> 59  
 cagaatagtg aagtcctaag gcctacagga tccatttgcg acaaagaaag aattcattca 60  
 gataaaaagc taaccgatag gcaaagattt tcaaattatt ttcttcttaa atatttttca 120  
 cattgttagt gaccaaagga gacctggatt tcaatcttga ctttggacct cactagctct 180  
 gcagtccttg gcaaatattt taatgtcctc cgaatccgtt tcctcatctg taagaaggga 240  
 tagtgtttta acttcacatg gttgagataa taatagtaaa caccttagtc ttgtgtttgc 300  
 cacagtggta gcacatagat attcaaagtt attattccta caaaaatagc gactgtactc 360

accccataat tgcatttttaa aagaaagacc agttttgcaa ttttccagga ttatctacat 420  
 gatagctttt tacacaatca gtat 444

<210> 60  
 <211> 240  
 <212> DNA  
 <213> Homo sapiens

<400> 60  
 cagactgtac gtcacacact cacttttgtt ttattacagg acaagtttac atagggtttg 60  
 aataccggga ggcagggatc attagggact atcttagagt ctatttatag taccctctaa 120  
 gttgtaatta aatttttttt tgtgatgaga tgtacacaac aatttagtat tttagccgtt 180  
 ttttaagtga cgattcaatg acatagtcac aatgtaggc aactatcacc attgttttca 240

<210> 61  
 <211> 598  
 <212> DNA  
 <213> Homo sapiens

<400> 61  
 gcctgggtaa gagtgagcgg agtgagacct gtctctaaaa ataaaaaacc ccaagctact 60  
 tatacattat atcatggtgt tatcaataag ggcaattaaa aaaaaaatcc aggaacattt 120  
 agcttgctgt tgccggctcag gagctcttgt aagggttcac ttaggatgtt acccaggggt 180  
 gaagtctgaa agctgcattg aactgaagga ttgacctccg agctcactta catgggttgc 240  
 gacaagggtgc tcaattgctc actatgtgga cctgtccatg gagctgttca tgacatgact 300  
 agtttaccct aatgagtgga tccagaagaa aacgtcatgc cacaattttt ttatgacca 360  
 gactgtacgt cacacactca cttttgtttt attacaggac aagtttacat aggggttgaa 420  
 taccgggagg cagggatcat tagggactat cttagagtct atttatagta ccctctaagt 480  
 tgtaattaaa tttttttttg tgatgagatg tacacaacaa tttagtattt tagccgtttt 540  
 taagtgtacg attcaatgac atagtcacaa tgttaggcaa ctatcaccat tgttttca 598

<210> 62  
 <211> 1430  
 <212> DNA  
 <213> Homo sapiens

<400> 62  
 agggtagaac ccatagccat ccatgttcat ctttgttttg aatataattg gctagaagat 60  
 gtacatatat ctatgtaact tcctctagca tcctccagta tggaggctgc attaagactg 120  
 gatgaaggag agggagagaa gggagaaaaca gagcagctgg acaagaggac aggtataggg 180  
 aataagggag aagccagtaa ggcaggaaaag accctccgtg acaaaggggc agggaacaga 240  
 actcaaacat ttaatggcag gtaaccagg ttagaatggg aaattgaaag gtgaatataa 300  
 agggagaatg gtgaaatgaa ttttctgaaa ttaattgctg tgtttatagt ttttagccat 360  
 gcatcggaat cacctcagga ctccactccc aatcaattat atatctgggg gaggaccaag 420  
 gcgttggtat ttttcagaag ctccactggt gattctgaca gcacagctag gattaagaaa 480  
 ctgatcaatg ggaacagcat gcctgttgca gaggagcttc cctgggaaat gtcacacaca 540  
 gaacatcaat cttctttccc cactcctgag atccctcatt ctttggcacc aggaacagtt 600  
 gcaattagta aaccctggtt ccctgctgtc tcacaaatcg caagagtcca acgtgtggat 660  
 ataaactttt gttcatggga ggatctttct cccagtggaa aagcaactgg gaaaagcagg 720  
 acacactgca cagtgactgc agtttcatcc aatgccacca cccatgcagg cataaataat 780  
 gaacatggat gggggagtct ggagctgctg aattgtaagg ctcataaatg tttaaacttt 840  
 ttccattaat aatatttctg ctttctgtgt atgtgtatgt agaagttctg tctttataat 900  
 tctcaccact ttgcatcata ctttccagga ggaagaaaga acacagaaat taaaattctc 960  
 acaaaggtta ccattaagct agagggaagac cacaccactg tgtgtccaca aagatacaga 1020  
 gccaggccgg gttcagccat gctggtcatc tgctctatat aatacaatta tttagagatg 1080  
 gtgggtagag aacaactaca gaaaaaaa aactgccaga aactagaatg tcatttttac 1140  
 acactcattt gtagaattcc tcccagtttt tactgaaggg aagtttaaaa tgattttcat 1200  
 ttgggggaaag aactgttttg agtttaccct ataagatggc cactaaaaact caccactttt 1260

```

catgattacc tagccatcct cagatcatct tcatgatttt cctggaaata acggaagagg 1320
ccctggggat gattttattg gtagagtggg aatgtattaa aattctctac ttccttgta 1380
catggtcttt cctccaccct acaaggtgtg tgcttgtaac tcaaatttcc 1430

```

```

<210> 63
<211> 3120
<212> DNA
<213> Homo sapiens

```

```

<400> 63
agggtagaac ccatagccat ccatgttcat ctttgttttg aatataattg gctagaagat 60
atacatatat ctatgtaact tcctctagca tcctccagta tggaggctgc attagactg 120
catgaaggag agggagagaa gggagaaaca gagcagctgg acaagaggac aggtataggg 180
aataagggag aagccagtaa ggcaggaaaag accctccgtg acaaaggggc agggaaacaga 240
actcaaacat ttaatggcag gtaaccaggg ttagaatggg aaattgaaag gtgaatataa 300
agggagaatg gtgaaatgaa ttttctgaaa ttaattgctg tgtttatagt ttttagccat 360
gcatcggaat cacctcagga ctccactccc aatcaattat atatctgggg gaggaccaag 420
gcgttggtat ttttcagaag ctccactggt gattctgaca gcacagctag gattaagaaa 480
ctgatcaatg ggaacagcat gcctgttgca gaggagcttc cctgggaaat gtcacacaca 540
gaacatcaat ctcccttccc cactcctgag atccctcatt ctttggcacc aggaacagtt 600
gcaattagta aaccctgggt ccctgctgtc tcacaaatcg caagagtcca acgtgtggat 660
ataaactttt gttcatggga ggatctttct cccagtggaa aagcaactgg gaaaagcagg 720
acacactgca cagtactgac agtttcatcc aatgccacca cccatgcagg cataaataat 780
gaacatggat gggggagtct ggagctgctg aattgtaagg ctcataaatg tttaaacctt 840
ttccattaat aatatttctg ctttctgtgt atgtgtatgt agaagttctg tctttataat 900
tctccacct ttgcatcata ctttccagga ggaagaaaga acacagaaat taaaattctc 960
acaaaggtta ccattaagct agaggaagac cacaccactg tgtgtccaca aagatacaga 1020
gccaggccgg gttcagccat gctggtcatc tgctctatat aatacaatta tttagagatg 1080
gtgggtagag aacaactaca gaaaaaaaaa aactgccaga aactagaatg tcatttttac 1140
acactcattt gtagaattcc tccagttttt tactgaaggg aagtttaaaa tgattttcat 1200
ttggggaaag aactgttttg agtttaccct ataagatggc cactaaaact caccactttt 1260
catgattacc tagccatcct cagatcatct tcatgatttt cctggaaata acggaagagg 1320
ccctggggat gattttattg gtagagtggg aatgtattaa aattctctac ttccttgta 1380
catggtcttt cctccaccct acaaggtgtg tgcttgtaac tcaaatttcc atttgagtaa 1440
ttagcaatta ttatttaaaa ctaacctgaa aataaaaaatt gtattcaatt cattcatagg 1500
gagcatctac ctattttatta ttaccacata ggtgatgtga ctctagaaac atcttggtat 1560
tcaaatagcc aattaaaata taaaatgtaa tgattttcta aagctactcg ttttcttct 1620
ctcatctcta tctactaatt ggataagtct attctccaaa cacagcaaag atgattgaca 1680
gaattctaaa aatacacaaa tttccctatt aaagaggggtg aatggatggt agcactgtat 1740
cagacacata atattaagga gatacctgct gtttaacatt aacattctgt gtagtttttg 1800
ttttgcatct ataacagaca tcaaaaagtg aataaaaaaca tgcgtgaaa ttacctaaat 1860
aataaattaa cttcctggat acaaggaaag tatttttagca agttcttttt aataaaagca 1920
aagaatgggg atgtaacaat tagagaaatg tggaggaaaa cataagaaag ttgggagagg 1980
gaggtcaaat ctcaattcca gttatcagta aagtgcacat tcttttctgt attctgtgag 2040
gctggagggtg ctgctgatag aatgtcctca catattctgt caatgccagg atgcaaacat 2100
cactaaataa accctcgtgt caaaatgtca catagtgtat atttatgctt cttgacaaca 2160
tcaatgaatg acaaggctgc ctctacagct gtgaaagggt gcataccaat tgttagaggt 2220
aacatagatg aaggacaaag tagttatcag aattcattta aactgtgcta ttaaattgga 2280
tatatttgca gtagccatca cagtaaatat tactgtgatt ctctgaaag aatattgcct 2340
aaaaaattta acgcagaaca tttcctaatt accatgcatt tctgaatctc tgctaccatg 2400
tagatattca tactgaaaga gaagtgtcat ttctttttct tctttaattt gtcattgttt 2460
cccagtgttc ttgcttcttg aattgaattt gaagtccatt ttttaagggt actgcaatta 2520
ttaacgaagt ctggcatgat aatgtctcag cagaatgtac acaacgtgac taagatagtt 2580
aacgccttag ttgttccaat tccatattgc ttctgttcc gaactacta atgagtaata 2640
agagatgtag cactaatcaa tgggacaaaa gcacgcatta tgaaaatact ctatcaactc 2700
catcagtaaa atttgtaagg taattataaa tttgttatca gacattttta attgttacag 2760
tataattaaa aggtattatt ttattgggga taaccttcca ctctgttaat cactactaaa 2820
cacttgatgc tcaaacacat tcaaagttac cttaatcagc gtgaatttca agatcctctt 2880
aaaaatgacc tctgatacta ccctgagatt actggagtgt aaatccttgc tttgcacttt 2940
ctggctgtgt gaacttaggc aattactaaa tttgctgttc ttgtattttc tctactgtta 3000

```

aatgaaatag aaatagcacc tacccctgtg aggtagtttg ataaaaggat taaatgagaa 3060  
aatctacata aagcatagtg cctaagatct gtcacacaat attttcaaga tccatgaaac 3120

<210> 64  
<211> 561  
<212> DNA  
<213> Homo sapiens

<400> 64  
ctccccagat atataattga ttgggagtgg ccaaggcgtg gtatttttca gaagctccac 60  
tggtagattct gacagcacag ctaggattaa gaaactgatc aatgggaaca gcatgcctgt 120  
tgagaggag cttccctggg aaatgtcaca cacagaacat caatcttctc tccccactcc 180  
tgagatccct cattcttttg caccaggaac agttgcaatt agtaaaccct ggttccctgc 240  
tgtctcacia atcgcaagag tccaacgtgt ggatataaac ttttgttcat gggaggatct 300  
ttctcccagt ggaaaagcaa ctgggaaaag caggacacac tgcacagtga ctgcagtttc 360  
atccaatgcc accacccatg caggcataaa taatgaacat ggatggggga gtctggagct 420  
gctgaattga ggaagaaaag acacagaaat taaaattctc acaaagggtta ccattaagct 480  
agaggaagac cacaccactg tgtgtccaca aagatacaga gccaggcccg gttcagccat 540  
gctggtcatc tgctctatat a 561

<210> 65  
<211> 632  
<212> DNA  
<213> Homo sapiens

<400> 65  
atgaattttc tgaaattaat tgctgtgttt atagttttta gccatgcac ggaatcacct 60  
caggactcca ctcccaatca attatatatc tgggggagga ccaaggcgtt ggtatttttc 120  
agaagctcca ctggtgattc tgacagcaca gctaggatta agaaactgat caatgggaac 180  
ggcatgcctg ttgcagagga gcttccctgg gaaatgtcac acacagaaca tcaatcttcc 240  
tccccactc ctgagatccc tcattctttg gcaccaggaa cagttgcaat tagtaaacc 300  
tggttccctg ctgtctcaca aatcgcaaga gtccaacgtg tggatataaa cttttgttca 360  
tgggaggatc tttctcccag tggaaaagca actgggaaaa gcaggacaca ctgcacagt 420  
actgcagttt catccaatgc caccacccat gcaggcataa ataatgaaca tggatggggg 480  
agtctggagc tgctgaattg aggaagaaaag aacacagaaa ttaaaattct caaaaaggtt 540  
accattaagc tagaggaaga ccacaccact gtgtgtccac aaagatacag agccaggccg 600  
ggttcagcca tgctggtcat ctgctctata ta 632

<210> 66  
<211> 398  
<212> DNA  
<213> Homo sapiens

<400> 66  
gcagcattaa atgaatgcct gttcattcaa gcatttcatt ttgttttaaa actttaaacc 60  
atttcatttc tatgacataa catcagtttt tctcatgcca agcctaaata tatatatatt 120  
cttggggatc tgtgactttt ctctataaga tacctctgtg tctttattcc aatgataatt 180  
ttagaaatta gcatacacct atttctgcat gatcttcata actatactgt aactttctat 240  
tagcataaaa tttcagtgcc acctttgtag cagtttttac aacagccttg gtactgcata 300  
attcatttta tgtattgaca agagatctga ttatagttat gttggcttgt ccaaagtcca 360  
caattagcca ggtgtggtgt ctcacacttg tgatccca 398

<210> 67  
<211> 2487  
<212> DNA  
<213> Homo sapiens



&lt;400&gt; 67

```

gatgatagga gttaagagag gactatagaa aactgggtct ctaagctgat gtgtcaagtc 60
acactgtcct ctgcttatcc taagcttacc ttgctcaaat ttcttttttt tttctttttc 120
tttgtttttg gtttttatct tttcttaaat ttcaaggata ttcttctctt tgtaaatgtc 180
acagagtatc atggctctgt cgccgaggct ggagtgaat ggtgcagtct caggctactg 240
caaccctgc cttccagggt caagcgattc tcctccctca gcctcccaag tagctgggat 300
tacaggcaca tgccatcatg cccggcta at ttttgtat tttggtagaga tgggggttca 360
ccatgttggc caggctgggtc tggaaactcct gacctcagggt gatttgccca cctcagcctc 420
ccaaagtgtc gggattacag gtgtgagcca ccgtgcccg cccaaccagg cttcttaaat 480
gaattctaag atagaaacaa caggagctgc caggactctc ttaagggctg aacctaggac 540
tgtcacagtg acatttctgc catattctgc tggtcacaag gcaagcccaa attcaaaagg 600
agagaaatag acctcttaga gtttccta at aaaaggta at ttcattaaaa atacaattca 660
taaattagcc ctatgtttac tactgtcttt tcagctcttt ttttattcca tgcattaatt 720
gattcgtcac cacttggatt gtgccaccaa tgtttctatg acatgatcta aaaaaaaaaa 780
aaaaaaaaaa aagggtcag tagttttcac ttaaaagaca aagaggccca ctgactaatt 840
acagatgtta gttaggattc atttacttta atatggtaga aagaatgcta tgataccact 900
ttagtgatga acaaaataag cttaatcaca tcctaggagc taagtattct gacattataa 960
tctcttctct cagagtccca tcacagcagt cttaggattc aagatctatt cttgggaaac 1020
attatagaac cagtgtgtca tgtacataca aatgagggaa aatataatgg ctttggtaat 1080
cctgtagtta tctttcttgt catatactct ttttttcatt tttaaaat ggggcaaat 1140
ttatataaca taaagttaac cattctgaag tgtacaattt aatggcattt aatacattca 1200
cagtgttcta cgactttttt ttttttgagg aaaagcatat ttttaggata atgtcaaaac 1260
agattaataa gatgctaata agatggccag acattcactc agaagtgttt tttgttttgt 1320
tttggtgaaa tggtagaga gatattgtgc cctacactta ggccactgca tcccgctta 1380
agtgccagga ttgtgtcagc aacaggatgg cctaaacaat ctcagtcttg tctcctgcca 1440
gcccctaaa tctccagaa ttgcaagaat aggcagggtg tgggtggctc cacctgta at 1500
cccaactctt tggaggcca aggcaggcgg atcacttgag gccaggagt tggagaccagc 1560
ctggccaaca tggcaaacc ccatctctac taaaaataca aaaattagcc aggcagggtg 1620
ggtgggtgat gccggtaatc tcagtttctt gggagggtcg ggtgggatga tcgcttgaa 1680
ctgggaggcg gaggctgcag tgagccgaga tcacgccact gcactccagc ctgggcgaca 1740
gaatgaggct ttgttctcaa aaaagtgtc aacacctgta accccaacac tttgggagcc 1800
acaggcattt ggattacttg agctcaggag tttgagacca gcctgggcaa catggtgaaa 1860
tcccactcc aaccgaaaat gcaaaaacta gccgggcata gtgttatgtg cctgtggtcc 1920
cagttacttg ggaagctgag atgggaggat cactagagcc cagggaagtca aggttgcagt 1980
gagccatgat tgcaccactg cactccacc tgggtaacag aacgagaccc tgtctcaaaa 2040
aaaagagact aaagcagcat taaatgaatg cctgttcatt caagcatttc attttgtttt 2100
aaaacttta accatttcat ttctatgaca taacatcagt ttttctcatg ccaagcctaa 2160
atatatatat attcttgggg atctgtgact tttctctata agatacctct gtgtctttat 2220
tccaatgata attttagaaa ttagcataca cctatttctg catgatcttc ataactatac 2280
tgtaactttc tattagcata aaatttcagt gccacctttg tagcagtttt tacaacagcc 2340
ttggtactgc ataattcatt ttatgtattg acaagagatc tgattatagt tatgttggct 2400
tgtccaaatg tcacaattag ccagggtgtg tgtctcacac ttgtgatccc aactcaggag 2460
gctgagggtg gaggattgct tgagccc 2487

```

&lt;210&gt; 68

&lt;211&gt; 1184

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 68

```

aaacactgca agcagaatgg gtcaaaggag agagaggcac acaatttcag gccaggccac 60
aaagccagaa aagatctgga aaatgttact ggaggaaata caaatcatc tggagtctaa 120
catttaaaga tagggtggta gattacgaga aaactaccat acagttcatg aggggtgggg 180
ggatcaaagc tataattaac aaactcggct aggagcaagc tacatattta ttctgcctgg 240
aaagtacttc tggcagtgtc ttatttctgt tgtccaaat tagtaatc atgaataact 300
ataacttcaa aaatgaagaa tatgattctg tgttttatac tttctacatg cacttatcct 360
attatgtaac tgaacgggtgt ggggaaaaga gtatatctaa tatgaaagaa aaaacctctg 420
ccctgagaa gtccaaatc taaatcgcc aactgtgaca aaagcaaggc 480
tttgaacatc tcagtatccc tggagtctgt cttctcgtca ttcacttatt gccactgtga 540
cataccacgt gacagtctgc atggccagaa atattaatta ctgtctgggg aagatgggat 600

```

```

tataatttaaat actccttttac tcctaatecct accattctcc aaaagtacaa tgaagagatg 660
ggcaggccttg gggatcctaata gctttcctac atcaaaaagg gacacattcc cattgccacc 720
ataagtttttg ctatatcttg gaagagggtca ctctttgaaa ttctatatgc aattttggga 780
aatagaatgg tgttcctttt tttttaccct cctaaaaaag aatgctgaag ttattttctc 840
tcttgccaaa aagagaagaa aaaaacgaga aggagaaatc agtttgagat ttgattgcgc 900
agacgtgaga acagaggcat tgccatacca tggcagatga ctcatccagc ctttaactct 960
ggctccaaac cctggcctct gcctgaggat tgggagaaag gcaatatgcc attctaactt 1020
gtagaccatt tatttaacaa ggttttcctg acttaaacct tctctttact gcaaagctgt 1080
acattggagt atgctgattt atgaaataat aaggactttt aaagtattca agagctggag 1140
attatgttta aaaagaaata actgttttat cagacctaaa gttt 1184

```

<210> 69  
 <211> 543  
 <212> DNA  
 <213> Homo sapiens

```

<400> 69
gtacattgaa aattgagacc tttggaggct tgtgcaagtt actgaatttg tgaagaaggt 60
ctcatttttc tttcttcttt ccagagtgtt tccttagtct cttcagctag tcatggctac 120
tactgttctg ccctcgtgag ccggtgtttg tctgccgggc ccgccttggg cacacagtga 180
aggctcattc cgtgtggttg cttgggtgag tctgcaccat ttggacacca agtctgtgtt 240
ggtcacacag tgaaggctca ttccgtgttg tggcttgggt gcgtctgcac catttgga 300
ccaagtctgt cttggtctca cagcgaaggc tcattccgtg tgggtggctt ggtgcgtctg 360
caccatttgg acatcaagtc tgtgttggtc tcacagagaa ggctcgttct gtgcagtggc 420
ttgggtgggt ctgcacacgt ggaacacacc acgtgaccag aggggcctga gtcaggccta 480
cctggagtag ctcagggaag ctgtaaagggt gagaagcaaa ggcttctcca tgtgtttctg 540
gga 543

```

<210> 70  
 <211> 560  
 <212> DNA  
 <213> Homo sapiens

```

<400> 70
gtacattgaa aattgagacc tttggaggct tgtgcaagtt actgaatttg tgaagaaggt 60
ctcatttttc tttcttcttt ccagagtgtt tccttagtct cttcagctag tcatggctac 120
tactgttctg ccctcgtgag ccggtgtttg tctgccgggc ccgccttggg cacacagtga 180
aggctcattc cgtgtggttg cttgggtgag tctgcaccat ttggacacca agtctgtgtt 240
ggtcacacag tgaaggctca ttccgtgttg tggcttgggt gcgtctgcac catttgga 300
ccaagtctgt cttggtctca cagcgaaggc tcattccgtg tgggtggctt ggtgcgtctg 360
caccatttgg acatcaagtc tgtgttggtc tcacagagaa ggctcgttct gtgcagtggc 420
ttgggtgggt ctgcacacgt ggaacacacc acgtgaccag aggggcctga gtcaggccta 480
cctggagtag ctcagggaag ctgtaaagggt gagaagcaaa ggcttctcca tgtgtttctg 540
ggacgcagaa cgcgcctaata 560

```

<210> 71  
 <211> 546  
 <212> DNA  
 <213> Homo sapiens

```

<400> 71
tccaccctga tgaggaggc agatatcaga aggacactgc acatgataat gggggtgctg 60
tgagggaata atgggggccc gggccttatg cagagggtgt gtcagggaag gccttgccc 120
ggggtgacat ctgagtttag agaggtcctg agaagaggcc tcaactcagg gctctggtca 180
tcagtcccag gagcccagcc ttccagtcac cccagcccgg gcaccagacg taagtatga 240
ggcttccaga tgattcctgc ccagctgtt cagggtctcc agcagagaag agctgtaccc 300
acagagccct gcttggcttc ttgacctgag gaatccatga cctgtttaca cccctaagtt 360
ctgtgatggg tcattataac aacagatcac cggaccatgg aaattacttc tcctcttcta 420

```

ctcttttaggc taactcgttt ttatataaac aactcatccc caccagaca cacacacaag 480  
 gccctagagc agcagtttcc gaacagtgcc gtatcccccg gggatctcta aaaaatactg 540  
 acaccc 546

<210> 72  
 <211> 676  
 <212> DNA  
 <213> Homo sapiens

<400> 72  
 tccaccctga tgagggaggc agatatcaga aggacactgc acatgataat gggggtgctg 60  
 tgaggggaata atgggggccc gggccttatg cagaggtggg gtcaggggaag gccttgccca 120  
 ggggtgacat ctgagtttag agaggtcctg agaagaggcc tcaactcaggt gctctgggtca 180  
 tcagtcccag gagcccagcc ttcgagtcac cccagcccgg gcaccagacg taagtgatga 240  
 ggcttccaga tgattcctgc cccagctgtt cagggctccc agcagagaag agctgtaccc 300  
 acagagccct gcttggtctt ttgacctgag gaatccatga ccctgttaca cccctaagtt 360  
 ctgtgatggg tcattataac aacagatcac cggaccatgg aaattacttc tcctcttcta 420  
 ctcttttaggc taactcgttt ttatataaac aactcatccc caccagaca cacacacaag 480  
 gccctagagc agcagtttcc gaacagtgcc gtatcccccg gggatctcta aaaaatactg 540  
 acaccagac aatagggggg aatcctaata atacctgacc aagactattc aagattgtca 600  
 ggctcatgaa aaacaaagaa aagtctgata ctacatcctg gatggaacat aacattaaaa 660  
 ctgagaaaat cggaata 676

<210> 73  
 <211> 451  
 <212> DNA  
 <213> Homo sapiens

<400> 73  
 aaaaagagaga aaagaaaacc caggaagttg cctgataatt ctttttcatg actactgaaa 60  
 agatggggagc ctgagagaga cctaacagga gtaataactt taaaaacaaa ttgtatggaa 120  
 ctgtacccat tctgaggatg actgatacag ttccagtatt caaaagaagt ataccttttc 180  
 cagcaaagca gatcttttgt ttacatagaa aacattttgt tatttgggaa gtttgcttaa 240  
 ctatttggtaa gtggctcttt gcaattaatg gattaccttt ttttaattgga aagaaaaataa 300  
 attcaaaatc cataagcaaa actttagtaa tttctgactc taggatacta gcactttcta 360  
 tttcttatgt gatttagact caacattttc cagttagtta agtacttaca tcctctgggt 420  
 gtatagctct gccatttgc ctcgtaatac a 451

<210> 74  
 <211> 453  
 <212> DNA  
 <213> Homo sapiens

<400> 74  
 tgccatcatat acattttctaa ggactcctct tctcagccat ctcaagctgc ttttatcata 60  
 agagatagga ccctgctcat tttgacaaga ggaaccatat acccagctgt gtcctccctc 120  
 catttgagta aacttccaaa acttgagcca ctaaaagatt ccattgctct gaccacctaa 180  
 cagttctccc aacatgggtc agagagcaga acagttctct aagctggtag ggggagagaa 240  
 gctctatcat gtttagagcag acatgaatgt gtaagggaga ataagcagag gagtgtgtgc 300  
 aataggcaaa ccatatctct gaagtgtgat gagtataccg aggtttctag ctggaaaaat 360  
 gttcaggggtg atgagataaa gaatatctca aagttttaat caataaatgc aaaccagcaa 420  
 aaatatcaat aaaaatattt taatgcctac aaa 453

<210> 75  
 <211> 481  
 <212> DNA  
 <213> Homo sapiens

```

<400> 75
tgccctcatat acatttctaa ggactcctct tctcagccat ctcaagctgc ttttatcata 60
agagatagga ccttgctcat ttgacaaga ggaaccatat acccagctgt gtcctccctc 120
catttgagta aacttccaaa acttgagcca ctaaaagatt ccattgctct gaccacctaa 180
cagttctccc aacatgggtc agagagcaga acagttctct aagctggtag ggggagagaa 240
gctctatcat gttagagcag acatgaatgt gtaagggaga ataagcagag gagtgtgtgc 300
aataggcaaa ccatatctct gaagtgtgat gagtataccg aggtttctag ctggaaaaat 360
gttcaggggtg atgagataaa gaatatctca aagttttaat caataaatgc aaaccagcaa 420
aaatatcaat aaaatatatt taatgcctac aaaaaaaaaa aaaaaaaaaa aattggcggc 480
c                                                                481

```

```

<210> 76
<211> 492
<212> DNA
<213> Homo sapiens

```

```

<400> 76
aattgaaata aatggtctga tgctctctaa aatcttactg ttacatttcc ataaagttaa 60
aaatgcaaac aaagtaagac ctccttgatt tggaacatgt cagaaagata agggaaacag 120
ttggtatgta ttctttattt agcaaagttg taatgcacag atttgacaaa tttaaagatt 180
ttttccctca gaatataatg taactttata ataagactag agatgtacct actgcatttg 240
atacccaaaa tgtatcattt ttttaatacct tcacttttat atgaaaattt actttaagta 300
atagtcattg aataatacac agtaattatt attgtcctca ttccttcttt agttttaaag 360
catttaataa ttaaacacag gaataccttt taaagtatta aaatttagta aaaatatttc 420
atctgcaaaa cagaatttgc atttaccaaa ctcaagcatt ataccttgca gtttgtaatt 480
gtttcactgt tt                                                                492

```

```

<210> 77
<211> 291
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (266)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (268) .. (269)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (273) .. (275)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (286) .. (287)
<223> a, c, g or t

```

```

<400> 77
tataaagtat tttcttatta cttctgctac aattttcaaa aggaattcct aggaaataca 60
cacatataaa acttaagaaa ccaaagaacc atttcaaaga atatttaatt tctttgttgg 120
tgctcaccaa atttctttat ataatttctg aaagaagtca aatgaaaaat taagttgata 180
atgcttatag aagtatttaa aggcatttat aaattaaata cattacataa ttatcaattg 240
aataaatggt tctacccatg gcagantnnt ttnnnntttt ttttggnacg g          291

```

<210> 78  
 <211> 870  
 <212> DNA  
 <213> Homo sapiens

<400> 78  
 ggccattca gtaaatatctt cttaagactc tactttgtgt ataacaaatt gattgtcaga 60  
 gcttattcat aaactatctt ataatgggtgc aaatcaactc acaaggcctt tgtctaataa 120  
 aagacagctg tgctgacagc tgcttgctct cttcaacacc tgatgtgctg gaagggtcca 180  
 cagtaaagca cttgaaacct tgaagtcagc cctactgtct ggataagaac ccattctctc 240  
 ttaaacctttt gaaatcaagg cttattaaat ttgattccta gaaacttaac gctagaccat 300  
 atttatcaag ttttaaattg ataaaattta aagtttttga gaagaaaagg gggaaaatat 360  
 gacaaaaaca ggcccttaagt tccaaggaag ccagtccttc agaaacccca tttcataatc 420  
 acccagttgg atgaaccaa gtagctttta cctcccaagg acagagcttc tcagaaaatg 480  
 gggcaaaaaa tctccagaca gccctactca gggtcctggg ctctgttctc atgttctgac 540  
 cccagaagg cctctaaatc actcaacttg gagactcgcc agtttttctt aatcagctgc 600  
 ctgaaagcgg ttcagagttc tgtcaacaag cccttgacag ctgggcttat taatgctggg 660  
 ccactcagag ccactgactca ggaacatggc ttgggggtcaa cactgaaaag cagaaatcac 720  
 agtgatgata atgggaattt tgtaggagga aacaggcttt tggaaactta tgcctttgtt 780  
 aggtttcttg atttgcagat ctcatgtgtt gggcctgctt taggaggaaa agctgggatt 840  
 cataacaatc taataaacct gacccaaacc 870

<210> 79  
 <211> 576  
 <212> DNA  
 <213> Homo sapiens

<400> 79  
 cccctcttgt atttatgatg gtttttcttt ctacttgatt aagaagaaag tacaaaaagg 60  
 tgcttattga agcattgtca atttgggaac aagattctca ggaatgaaat atttgtgact 120  
 gttctcagca tgaaaggtag agagaagatt acttgatcat tatgtgctta tttgaaagtc 180  
 actttcctcc catgcctcaa catctcactt tttagaatgc aactgaatg cctggaatct 240  
 ctttgtttta gaaaagacaa agttgcattg ttaaaaaaat aatatctctg gggaaaaggc 300  
 tcagtgctca acggttactc tataaagaga tttaaaaaga aagtcaaggg aagataagga 360  
 aggtagcatg gctaacaacc tcgtgggttc tactagcatt atttcatgca aaatgtcatt 420  
 aacctttata taggagagaa taagtctggg aatcaactgg catctaaact ccatttcttg 480  
 ttttgctact gcttcctcgg ttgactatgg aaaagacatt tgggtctttg aaaccttgcc 540  
 cccaacaact ataacatgac ttaaatcat ttatgc 576

<210> 80  
 <211> 905  
 <212> DNA  
 <213> Homo sapiens

<400> 80  
 ctgacccctca gttttctgtc aaagaacttc tgctcatatc caattcctat aaagtgggag 60  
 tactctaacc tgatattacc aacattaggt tagattttat cctaattgta ggtattagaa 120  
 tatatcctaa cattactagt tcttcaactc tactaaagtg catgttagat ttcttcaatg 180  
 atttttttaa attggcatgg ttgatacaaa gcacccctct atccaaatac tgtgctcagt 240  
 cgcctaagtg atagtttaga accatactat ctctctgtgg aaagaatgct ggggctcctg 300  
 gctacacact ttgagtagaa gggaaaatac ccctcttgta tttatgatgg tttttctttc 360  
 tacttgatta agaagaaagt acaaaaagggt gcttattgaa gcattgtcaa tttgggaaca 420  
 agattctcag gaatgaaata tttgtgactg ttctcagcat gaaaggtaga gagaagatta 480  
 cttgatcatt atgtgcttat ttgaaagtca ctttctctcc atgcctcaac atctcacttt 540  
 ttagaatgca cactgaatgc ctggaatctc tttgttttag aaaagacaaa gttgcattgt 600  
 taaaaaata atatttctgg ggaagggtc cagtgtcaa cggttactct ataaagagat 660  
 ttaaaaagaa agtcaaggga agataaggaa ggtagcatgg ctaacaacct cgtgggttct 720

```

actagcatta tttcatgcaa aatgtcatta acctttatat aggagagaat aagtctggga 780
atcaactggc atctaaactc catctctgtg tttgtcactg cttccctggg tgactatgga 840
aaagacattt ggtcttttga aaccttgccc ccaacaacta taacatgact taaattcatt 900
tatgc                                           905

```

```

<210> 81
<211> 622
<212> DNA
<213> Homo sapiens

```

```

<400> 81
tgagatttct ttatagtcct ctgtaggtaa ggggttgattt tgaaattttg gctgcttggtg 60
tttaagattt gattctcatt ctcaaacatt tacatgtata attttgagaa cacattttat 120
ttaacacaaat atagtaaaca gaatgaatca taactatgaa cgttttccag aagcatttct 180
tagataaatt attttataaa aagaaaaaca acacaacatc taagatttag agtgggaaga 240
tatacttgta gttaacttcc ttgttgattt aaaaaatata tttggattta ttttggcagg 300
gtgggggaat catctaataa aaaattttaa gcaaacttca ttttttctaa ccagagtga 360
gagacagggg gagaatcaaa tatgtgtgag ctccctctgt tgctctgtaa cagcatttta 420
ttcatgattt gtgatgtgat aaggaaattt ttgctaattg gtcaaaacaa attcaatttt 480
cctttaaggc aatttataac tttattttaa tggaatataa aggaagaacc ccctatatgt 540
aaaattctga ccttcaaatt tatgttaata tttttaaatt attaaaacat taaaaatgca 600
tctatttctc accactaaga gg                                           622

```

```

<210> 82
<211> 1079
<212> DNA
<213> Homo sapiens

```

```

<400> 82
gaaagatcta aatatccctg cacctcaagt agtacctatc atatgtagac accatataaa 60
tatttgtgtt gaatttaata aactcaggct gaacctactc tgcatatatt gtgatgactt 120
tcagattaac ctttagacac attcacaggg ttactctctt ggaagatctc tagggatcct 180
cattttattt ggtcttttgg gttaactgtat gttctgctgt tgggtgtttt tactctttga 240
gtattgtttt tcttccttta caaagatagt gtacatgttt attggttggg gggagtcaaa 300
tcatatcaaa tgatagaaaa tgaaagccat ctccctagtt ctattcccca gaagagacca 360
ctgataagag tgtagtgtgt atcttttaag actattttca gcacatacat acatacgtat 420
tttgcacata tatggcataa ttaatatatt gttctaaaat ttacattttc atctatctag 480
ctagctagct atatctcacg tatttccaag tgcagtattt tgggaagttt aaatagtcag 540
attagggcca tgttttggtat tgtgactttg ttcctgacct gggtagagtg ctgtgcttag 600
ctgacatatt ttacaatcca gtgatcactc ctgacattat ttgcagagtg ctcttaccag 660
cagaggagaa agagctactg tcccaaaaga ttggagcaaa tagccctgag ggaagtggaa 720
aatgtctttg gagtggtatt tcttttatct taaaatttag tgcagatctt gcattcaaag 780
acatcatggt atatctgtgt ttgtttcctt tgtttttaca aggagtttct cccaaaaaac 840
tgaacctgaa gtaatggtca ctccaggaat ttatgtcttg tttatcctat ggcttcaagt 900
accttcagtg tattgccaag tactgctcat gtacactcca gctgagaaat acagtgttag 960
gtttccaaca aaagtaccta aggatttccc acagttcact taagaaggat gcaaaagatg 1020
ttactggaat aatcatagtc gcagttagct gtagaataaa ggacaggacc cggtaggt 1079

```

```

<210> 83
<211> 331
<212> DNA
<213> Homo sapiens

```

```

<400> 83
gtcgggtattc atctcccctg ttgggaaagg tttaatctca tgggttattt cccaaaacct 60
gctttgttgc tctctcattt gtaaagcatg caatatcagc tggagagagg aaggcactgg 120
aaagaatgct ttggtcactt tacatcagtt ttaaagtagt ggcaataaag agaattgccta 180
ttcagggcat ttattggcac ttccatgggg ggtaggtcct gaagtttctt gtggcagggtg 240

```

agtgaaaggc ctggaagaa ggccaaggat gaaattgatg tggagaagag gatctggctg 300  
acttttcctt gagaattcta agggatattt c 331

<210> 84  
<211> 437  
<212> DNA  
<213> Homo sapiens

<220>  
<221> unsure  
<222> (362)  
<223> a, c, g or t

<220>  
<221> unsure  
<222> (399)  
<223> a, c, g or t

<400> 84  
gtattttcca gttattcact ttcaactttat aaacatgaca cttttgttaa gattataaaa 60  
atgaattttg actgtgcac tgcaatttta gacattttcg ttatgattgg taatagaaca 120  
ataaaatgtc tagccttgta aagagttaag ttactttaca agacagacat catctaattt 180  
gcaacaagga ttaaagcccc actctgaatt aaaatttctg ctggaaagat ttgaggcact 240  
ggcaaaactgt caaaggtagt attgtgcgag agtaaccagc ttttgcaaga cctgtgaaag 300  
taaattctctt cggccagagt ttcttaggtg tgtgaactta tgatgatcag gaatacattg 360  
gnggattccc actagagaaa cgcaccatta atagttaang ctttcacttt ctggtttcac 420  
ttgttgccct cataaat 437

<210> 85  
<211> 526  
<212> DNA  
<213> Homo sapiens

<400> 85  
gcagtatttc actgccacct gccaccaggg tttaaacagg ggagtagcag ggagctatta 60  
gggaggaggg gattgagga tttttactaa ggcagaggca ggtgatagat ttgagatttg 120  
caaagtcaga acttgagacc cagttgagga atagcattac aatgtagca gcgttgtag 180  
gcgagcacia gacttcacag gtgatgctag ttcacccaaa tgtgaaccac cattaactgg 240  
ttttcctttt gaatgtgatg gttccaagag gtttgagtc agtggagatg tgtgaaaggg 300  
cttgagggtg gaaatctggg taagaatgcc aagggcattc ctggtagatt aaaatggtaa 360  
agcaagcaac agatctggaa aagaactagg agaaactggt agtattttct ggggtgtcaac 420  
gtagggaaag gcgttctaag cgttttcctt ataaataatg aaaaaatggt aaaaagccaa 480  
aaaaaaagtg gggggagact gaaaattgag attatataga aaattg 526

<210> 86  
<211> 440  
<212> DNA  
<213> Homo sapiens

<220>  
<221> unsure  
<222> (144)..(169)  
<223> a, c, g or t

<220>  
<221> unsure  
<222> (178)..(179)  
<223> a, c, g or t

<220>  
 <221> unsure  
 <222> (181)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (218)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (403)  
 <223> a, c, g or t

<400> 86  
 gctcagagtgt taaagtaata agtaacaaaa taattagaaa agttagaaaa ataatagacta 60  
 tttggcaaaag atattttctca tacaatgaga aatatctttg tccaattagt ttaaaatctg 120  
 atgtagagaa actgtatatc tatnnnnnnn nnnnnnnnnn nnnnnnnnnng atacttanna 180  
 nactacttta gtttttagag tagttttagt ttccggcnaa aatgagcagg tacagagaac 240  
 atttttgttt acaatctaga caaatgtatc cattgtgtac aaattcattg aaaataatgt 300  
 tatattatgt tatatgtgta ctcaaatact ctgggttgta attcagcaaa aacttggttt 360  
 ttaacaagta gcttcacatc catttttgtt attttcaata aanacaaatt cttgtcatta 420  
 tgcaacaagg ttataataaa 440

<210> 87  
 <211> 95  
 <212> DNA  
 <213> Homo sapiens

<400> 87  
 acaggcgtga gtgccactgt gcctgtctca ttcctctctc attattagct ggaatacttc 60  
 cagaaagaga catttccctt actgactgaa acaat 95

<210> 88  
 <211> 416  
 <212> DNA  
 <213> Homo sapiens

<400> 88  
 cgcttggtgtg gggcaggggg tatatgggaa ccccatataa ttttactgtg aatccaaaaac 60  
 tgctctaaga aagtctggct tttttttttt ttttctgaga cagagtttca ctctgtctca 120  
 ggctggagtg tgatggcacg atctaggctc actgcaccct ccacttccgg gtccaagcga 180  
 ctctcggtgc tcagcctctc aaagtgctag gattacaggc gtgagtgccg ctgtgcctgt 240  
 ctcatccctt ctccattatt agctggaata ctccagaaa gagacatttc ccttactgac 300  
 tgaaacaatt ttttgaattt ttttatttta tatatgagag ggagttttgc tcttggtgct 360  
 taggctggag tgcaatggcg tgatctcggt gcactgaagt ctccgcctcc tggatt 416

<210> 89  
 <211> 270  
 <212> DNA  
 <213> Homo sapiens

<400> 89  
 cttccagaca gctggccagt tatgttactg aaacagatat gtttctgaaa catacctatg 60  
 gctatattta tacctgggtt ttaaaacttt gactcttatt ggcttcttct ggtacatttt 120  
 ttatttgtaa tttgtcagaa tttcagtatc tgagagtata cattctgaat tttatagttt 180



tagatcatgc aattacactt atttttcctt gcttttgaaa aagtaaagt gcttcctatt 240  
 ttcttaatga tcagttatct tttttgttg 270

<210> 90  
 <211> 148  
 <212> DNA  
 <213> Homo sapiens

<400> 90  
 gatggatggt gttacttagc ttgaagaagt acattaaact gcactgggtct ttggcaacac 60  
 gtcccacgtg ccatgctagg catgcaatgg attctgatct tttattgtac aagtgggtgta 120  
 aattctgatt catgacgata tgttggtg 148

<210> 91  
 <211> 853  
 <212> DNA  
 <213> Homo sapiens

<400> 91  
 acaaatgtca tcactgatag aaagggttcag taacttactc aagaccacaa aattaggtct 60  
 ttttgaaaat taaaaaaat tcaaaccag gatataaact cagggtttatt ttattcccaa 120  
 gcctattaat ttcaagttac agttattatc aaactgtatg tttcttaagg caggatctga 180  
 gttgtgtaat catttatccc catagtagct tgcccttaag aggtacttag tacatatattt 240  
 ttgatgaatg atgtgtaca aataatggtg tcctgtataa taggttatgg tttaaaaata 300  
 gaggaagtct ataggactct tagaaagtac ttcagatata tgagaagact aagattgaga 360  
 acttcttgga ggtctaattc taaattaaac ttccaagttg ggatgcaaaa ataaggagga 420  
 tgtttggaaa catcccatga tttatatatta cataagcttc ataaggagaa tgaatacaga 480  
 ctgacctttg cttgagagcc atgtgatggc cagagaaagc ggcatgggtt aacagccaaa 540  
 aggaaatcag cattatacta aagcagtgtt tctcaaagtg tgttccaggg aaggatatac 600  
 ccaagaaaca atatccatca agagacttct ataaaaagag gagataattt aatacatata 660  
 gaagcatgac atagtccac ggatccctga acaatacagg ggttgggggc attgacctt 720  
 gtgcagacaa aaatctgcat gtaactttt tttgagacag agtctctctc tgcactccag 780  
 cccgggcaca gagcaagacc ctgagacctt gtcagaaaga aataaagaga gaaagagaga 840  
 ggagagggag agg 853

<210> 92  
 <211> 801  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (553)..(676)  
 <223> a, c, g or t

<400> 92  
 cggagagaga ctctgtctca aaaataaagt aaatgaaata aatcttacgt ttaaaagccc 60  
 tttttcatgg attttcctca gagactttgc agaaaccaag gggtcagtta tttatcagtt 120  
 ggaacatttt ggactttatg atcacaaatg ctttttaatc tgtgtaactt catctacaga 180  
 aaaccatggt ccattaagac tagagcgatc cctctatctc ttcatgccag gcttctacca 240  
 gggagataac ctgtacacat tactcacgag catagtgcgt aatccacata gggtaatgtc 300  
 tgcagatttg agtatgtgtt ccctaattct ttatctagca tgtagagtat aaataacaca 360  
 atactggatg cttttatgga tgaacaagga ataataccta gcacctttct tctagaagtt 420  
 tatagtatga agagagaaga taagatgcat ctgagaaact agattaaact tgacattggt 480  
 tgatcaagag ccacgtgagc aataccaaca tgggttgaga cagaggagaa atccatcgtg 540  
 gtgaatctag agnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600  
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 660  
 nnnnnnnnnn nnnnnnccac ctctgccccca ggtacacacg gttggtataa ccaaaagtat 720

ctccaggtgt tgccaaatct gatctagagc aagaaaaaac atggaacatg aaaacagtgt 780  
gtttaaaata aagccagaga g 801

<210> 93

<211> 280

<212> DNA

<213> Homo sapiens

<400> 93

gagaacagga ccccatgctg gcacagccct ggcttaacgc gggggtggcc ccagaagcgg 60  
gtgggcgagg ctgggcagca gggattggct gaaatcatat gcagagccca agaggcaggg 120  
gaaaggcggc aatttcaggg tccctttgtt cgccaggtaac ctggggccca gcccgggcgg 180  
caggagggac tcagccctc gcccaggcag gaagggtccc aagcagaggc ccctccctca 240  
ggcactcccc agcccacacc tgcagcactg ggccaagact 280

<210> 94

<211> 829

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (784)

<223> a, c, g or t

<400> 94

gaacaggacc cccatgctgc acagccctgg cttaacgcgg ggggtggcccc agaagcgggt 60  
gggcgaggct gggcagcagg gattggctga aatcatatgc agagcccaag aggcagggga 120  
aaggcggcaa ttccagggtc cctttgttcg ccaggtaacct ggggcccagc ccgggcggca 180  
ggaggggactc agcccctcgc ccaggcagga aggggtccca gcagaggccc ctccctcagg 240  
cactccccag cccacacctg cagcactggg accaagacta ataaaaacacc cgcctcacgg 300  
aagacagctt tatcttgttg atcggaagtc tgccagccca atttatgatg gaacataaga 360  
tctctaaatc tgaatttaca ctctgtagcg taacgagagg tcaataagat taaacggggg 420  
ctcaggagag gaccagcgtc aggtcactg cgagggtgctg cacagaaaac ccacagccag 480  
agcccctggg ccagcccag gcaagaccag aaaaggaggg ggcaagggtgg agaccagcct 540  
ggggctcccc ggaagcccac gggatggagg cgggagagcc aggaggcctg gggcaaccct 600  
gggacgggtc ctggatcgag gagagcaggg ggggtgatgag ggttccctca gggctgggga 660  
gccttctcct ggtctcagac ccaccccctc tcagctccca agccctgggt gccctgggt 720  
ctgaggacag tgggaatctt ccctgaggca ggttcaagga cagagctctg accctgtgcc 780  
aggnctgctt tgggtgccta tgaactcggg tctggctcag agcagtcct 829

<210> 95

<211> 170

<212> DNA

<213> Homo sapiens

<400> 95

gataaccagt aggcagcagg ctactcctgg ttcgtaacat aactccacca gtgtattctg 60  
gcagccgggc ttacttaagg gagagcacag acattccctg ctcaaaaaa aaactgctaa 120  
acgtgactcc ggtagcctcc atgctctctg caagaataaa atccttgaag 170

<210> 96

<211> 259

<212> DNA

<213> Homo sapiens

<400> 96

```

caaatTTGaa atcttaaaat ttaagaaact agtggaggaa ttggatagta catgatttca 60
aaaacatgaa aactgaggac attaaatgtg caagggttag aagtttgctg catgcaaagg 120
ggaaagtgaa gatagcattt ttccacatag ttccagaagt ccagttgctg aggttaataca 180
atgaaagttg tagcatcaaa ggtttaacat aaaacaactt cctaaatca gccagggtgca 240
gaggctcacg ccggtaatc                                     259

```

<210> 97  
 <211> 392  
 <212> DNA  
 <213> Homo sapiens

```

<400> 97
atgggtttgg tcattcaact ttaagtttat gagatatatc catggtgaat tttgtagctg 60
tggtttgatg atttttacat tatatagtat tacattccat ggatagttct cagtagataa 120
tctcctatt gtttaacatt tgcgttgctt ctcattttga cctattttta acaggcctct 180
ttaaataatac acttctctgt agtgtatgct agaaatggag tggctggaat aaaagtggct 240
gaatcatctt caactctagt aagatgtcaa actgttttct aaaagtgtt tattttaact 300
attatatgtc aattttgaac agctctttca cttactagca atttattatc agcaacactt 360
gttattgtca gacttttaag ttttcattca ct                                     392

```

<210> 98  
 <211> 863  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (395)..(627)  
 <223> a, c, g or t

```

<400> 98
atgggtttgg tcattcaact ttaagtttat gagatatatc catggtgaat tttgtagctg 60
tggtttgatg atttttacat tatatagtat tacattccat ggatagttct cagtagataa 120
tctcctatt gtttaacatt tgcgttgctt ctcattttga cctattttta acaggcctct 180
ttaaataatac acttctctgt agtgtatgct agaaatggag tggctggaat aaaagtggct 240
gaatcatctt caactctagt aagatgtcaa actgttttct aaaagtgtt tattttaact 300
attatatgtc aattttgaac agctctttca cttactagca atttattatc agcaacactt 360
gttattgtca gacttttaag ttttcattca ctggnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
nnnnnnnnnn nnnnnnnnnn nnnnnnnngtg atcatattta tgttttgctc atttaaaaaa 660
acctgggttaa atatttcaca aatcgacatt atgatataat atcttccaaa attttaataa 720
ttttgtcttt ttccacattt tagtcttttag ctcagctgga attcatttct gtgtgtggtg 780
tgagataagt ctttttcattg tttttcccta tgaaataaat tatttccttc tgtattgcag 840
gtagctgaga ccaatgatag ctg                                     863

```

<210> 99  
 <211> 563  
 <212> DNA  
 <213> Homo sapiens

```

<400> 99
ggtacatctt ggctgtggat ggaaatttga catacttttt attttcttta cctgttacat 60
atcaaactctt aggatgtatt acttccaagg cgggttaaact tattcaagat tgactgagtc 120
tctatttttc cttaaattta ctagaagtga ggctccaaga actacagaaa atagaaggaa 180
agtctccatt gagccatgaa ctgtgagcac ctggcattta agcatgaaga gtagggcttc 240
tatggtaggg actggagtag gcagcattcc aggaaaggat ctcagaggtc agaaacaata 300

```

```

gattatcagt taaatacttc tggaccaaag aagaccttga aatcctggct cggatgatgta 360
tacatattta atacacaatg agaagctcct tgagggtgag tgagagtgat agggctgaag 420
aacaggagac agaagacaat tcaaatgtcc ttacacagaa gactgattat atagataatg 480
gtacattcat ataaacatga tatatttact aattaaaaaca taccaacaca cacaacacct 540
cgagccgcta gtctcgagtc tag
563

```

&lt;210&gt; 100

&lt;211&gt; 667

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 100

```

cggcaagctc taatgggaag gatacgccct gatcacacac tcctatttca aaggggtcca 60
gtaccagcac ctctcaccag tggctacac tattatacaa ctttagaaga actctggaaa 120
agttttgatc tttgtgaaga ctattttaaa cctccatttg gaccatatcc tgaaaagagt 180
gggaaggatt ccttggtttc catgaaatgt tcattgtttc ggttctgtcc gtgggtcaaaa 240
gaattgcctt tccagcctcc ggaggggagc atttcttcac acctaggatc aggagccagt 300
gacagtgaga ccgaagagac ccggaagca ctacctatac aatcattttc acatgaaaaa 360
gagctcacc aacacagaca aactcggtc ccagtcatac gtcgcccagg ttccaacgtc 420
aaacccaccc tccctccaat ccctcagggc cgcaggtaga ctagcacttg atgtctgac 480
ctaacatgga aaacctgtc tgctgatgtc gaattccttg ccttacctgg ccatgggtcc 540
agctgtttct cactcaaccc attaccacg gaagaatgtg ttacctgcc ttaattctat 600
cagccagttt ctcttgatgat tctttggctg gtgtctttta gttttttaat taaaaaattg 660
tttctta
667

```

&lt;210&gt; 101

&lt;211&gt; 3734

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 101

```

gagtatacat ctttttcagc actgcacaac acctattcca aaattgacca catagttgga 60
agttaaagctc tcctcagcaa atgtaaaaga acagaaatga taacaaactg tctctcagac 120
cacagtgcaa tcaaaactaga actcaggatt aagaaactca ctcaaaactg ctcaactaca 180
tggaactga acaacctgct cctgaatgac tactgtgtac ataacaaaat gaaggcagaa 240
ataaagatgt tctttgaaac caacgagaac aaagacacaa cataccagaa tctctgggac 300
acattcaaag cagtgtgtag agggaaactt atagcactaa atgtccacaa gagaaagcag 360
gaaagatcca aaattgacac cctaatatct caattaaaag aactagaaaa gcaagagcaa 420
acaattcaa aagctagcag aaggcaagaa ataacctaaa catcagagca gaagtgaagg 480
aaatagagac acaaaaaacc ttcaaaagaa ttaatgaatc caggaaactg ttttttgaaa 540
ggatcagcaa aattgataga ccgctagcaa gactaataaa gaagaaacga gagaagaatc 600
aaatagacgc aataacacac atgataaagg ggatatcacc acgatccac agaaatacaa 660
actaccatca gagaatacta taaacacttc tatgcaaata aactagaaaa tctagaagaa 720
atggataaat tcctggacac atacaccctc ccaagactaa accaggaaga agctgaatcc 780
ctgaatagac caataacaga ttctgaaatt gcggaataa ttaatagcct accaaccaaa 840
aaaagtcag gaccagatgg attcacacct aaattctacc agaggataaa agaggagctg 900
gtaccattcc ttctgaaatt attccaatca ataacaaaag agggaaatcct ccctaattca 960
ttttatgaag ccaacatcat cctgatacta aagcctggca gagacacaac aaaaaaaga 1020
gaatttagac caatatccat gatgatcatc gatgcaaaaa tcctcagtaa aatactggca 1080
aaccaaattc agcagcacct cataaagctt atccaccacg atcaagtttg cttcatccct 1140
gggatgaaag gctggttcaa catacgcaaa tcaataaagg taatccatca tataaacaga 1200
acaaagaca aaaaccacat gattatctca atacatgcag aaaaggcctt tgacaaaatt 1260
caacagccct tcatgctaaa aactgtcaat aaactagtta ttgatgggac atatctcaaa 1320
ataataagag ctatttatga caaaccaca gccaatatca tactgaatgg gcaaaaactg 1380
gaagcattcc ctttgagaac tggataaaga cagggatgcc ctctctcacc actcctatc 1440
aacatagtgt tggaagtctt ggccagggca atcaggcagg agaaagaaat aaaaggtatt 1500
caattaggaa aagagaaagt caaactgtcc ctgtttgcag atgacatgat tttatatata 1560
gaaaacccca tcgtctcagc ccaaaatctc cttaagctga tgagcagctt cagcaaggctc 1620
tcaggatata aaatcaatgt gcaaaaatca caagcattcc tatacaccaa taacagacaa 1680

```

```

acagagagtc aaatgagtga actcccatc gcaattgctt caaagagaat aaaataccta 1740
ggaatccaac ttacaaggga tgtgaaggac ctcttcaagg agaactacaa accactgctc 1800
aacaaaaataa aagaggacac aaacaaatgg aagaacattc catgctcatg gatagggaaga 1860
atcaatattg tgaaaatggc cataatgcc aaggtaattt atagattcaa tgccatcccc 1920
atcaagctac caatgacttt ctccacagaa ttggaaaaaa ctactttaa gttcatatgg 1980
aaccaaaaaa gagcccgcat tgccaagaca atcctaagcc aaaagaacaa agctggaggc 2040
atcacactac ctgacttcaa actatactac aaggctacag taacccaaac agcatgggtac 2100
tggtaccaa acagagatat agaccaatgg aacaggatag agcccttgga attaatacca 2160
cacatctaca accatctgat ctttgacaaa cctgacaaaa acaagctatg gggaaaggat 2220
tccctattta ataaatgggtg ctgggaaaaa tggctagcca tatgtagaaa gctgaaactg 2280
aatctcttcc ttacacctta taaaaaatt aattcaagat ggattaaaga cttaaatgtt 2340
agaccgaaaa ccataaaaat cctagaaaaa aacctaggca ataccattca agacataggc 2400
gtgggcaagg acttcatgac taaaacacca aaagcaatgg caacaaaagc caaaattgac 2460
aaatgggata taattaaact aaagagcttc tgcacagcaa aagaaactac catcatagt 2520
aacaggcaac ctacagaatg ggagaaaatt tttaaatct accatctga caaaggacta 2580
atatccagaa tctacaaaga acttaacaaa atttacaaga aaaaaatcaa acaaccccat 2640
caaaaattgg gcaaaggata tgaacagaca cttctcaaaa gaagacattt atgcagtcaa 2700
cagacacatg aaaacatgct catcattact ggccatcaga gaaatgcaaa tcaaaaccac 2760
aatgagatac catttcacac cagttagaat ggctagtatt aaaaagtcag gaaacaacag 2820
gggtccagta ccagcacctc tcaccagtgg tctacactat tatacaact tagaagaact 2880
ctggaaaagt tttgatcttt gtgaagacta ttttaacct ccatttggac catatcctga 2940
aaagagtggg aaggattcct tggtttccat gaaatgttca ttgtttcggg tctgtccgtg 3000
gtcaaaagaa ttgcctttcc agcctccgga ggggagcatt tcttcacacc taggatcagg 3060
agccagtga acgtgagaccg aagagaccgg gaaagcacta cctatacaat cattttcaca 3120
tgaaaaagag tctcaccaac acagacaaca ctcggtccca gtcacagtc gccagggttc 3180
caacgtcaaa ccacccctcc ctccaatccc tcaggggcgc aggtagacta gcacttgatg 3240
tctgatccta acatggaaaa cctgctctgc tgatgtcgaa ttccttgctt tacctggcca 3300
tggttccagc tgtttctcac tcaaccatt acccacggaa gaatgtgttt acctgcctta 3360
attctatcag ccagtttctc ttgtgattct ttggctgggt tcttttagtt ttttaattaa 3420
aaaattgttt cttacaaaaa aaacggtatt cctggaggcg attactggta cgcaatctag 3480
gcaatgctcg attcagccca ctacaccctg ccctgtggg ccccgagtgg ggacctttgc 3540
ggccccccct tttgtgggg gggttaacgca tagccccctt taaaggatac cgttcccccc 3600
ggcctgcagg ttggcacaac aggtttaatc ctagtaaagg ggggagatcc cgctcttccc 3660
acaatcgagg ggccccgaac aaaaataacc tgtaatcaat gaggcccgac aaggccagaa 3720
cacacaagag cggc 3734

```

<210> 102  
 <211> 353  
 <212> DNA  
 <213> Homo sapiens

```

<400> 102
gtggatggca gcacagattt agtgtctggt tttgtcagac ccatagctgg tgaggaaaac 60
agttaaccgg aatggaaatg aaagaaagtg gttcgagggt acagcaaagc aaagaacaag 120
caaagcatag ctggtaacag aaaaaaagac tgaggaatga taaacaatgg acgtaagaac 180
tccaaggcca gcaataggag gccatctaga ggaagcggcc agcacacctc gatagccata 240
gggcagcgct gtttgtgcca ggaggaatgatgtacaga tgggtgacact aggacatgaa 300
gaaagggtta ggcgatgttg agaaacctca ttttaagaag acacacagtc gga 353

```

<210> 103  
 <211> 484  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (8)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (11)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (285)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (427)  
 <223> a, c, g or t

<400> 103  
 cggctcgnag ntgttacagc tatggataat agtggttatct aagctgagtg tattgaatac 60  
 tttcatctag atgtcttaaa aggacttttag taggatataa aaagaaaata aaaaggaatt 120  
 ttaaaaatca aattaaggca aacaatctct acacattttc ccccttgat ctatgccttt 180  
 acatttaggt tataaagtca gccccctcc ccagtccac ggattagcca actatctttc 240  
 cgtttttgat tgtntggtag tgagtactgg agaataaatg ctgtntttta tattatgtat 300  
 ttttttcata actgaaattg gcgttagaag attggattaa aaaacatgag acctaactt 360  
 ttagataatt gtaaatatag tgtaagcagg aagatattta ctattttcaa taataaagac 420  
 ataaatnact tgttttctgt attttaagaa acttttgctg gtatttgata ttttaaaaaa 480  
 taag 484

<210> 104  
 <211> 119  
 <212> DNA  
 <213> Homo sapiens

<400> 104  
 gagttttact cgcacccgat gagaaagggtg tgtgtacctg ctttcatgac gatagagagc 60  
 agacaacttc tttctggagt ttcagcttgc ttccaacagt gaaggaggaa ctgaaattt 119

<210> 105  
 <211> 290  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (19)  
 <223> a, c, g or t

<400> 105  
 gttaatatc tcataagcna tacacacctc gaagccatca atgacaacct tttcttgctg 60  
 aatagaacag tgattgatgt catgaagaca attttatctc cttttgcctt ccataatttg 120  
 taccagtgtt tgtcacgtgg ttgttgaata aatgaataaa gaatgagaaa accagaagct 180  
 ctgatacata atcataatga taattatttc aatgcacaac tacgggtggg gctgaactag 240  
 aatctatatt ttctgaaact ggctcctcta ggatctacta atgattttaa 290

<210> 106  
 <211> 1645  
 <212> DNA  
 <213> Homo sapiens

<220>

&lt;221&gt; unsure

&lt;222&gt; (60)

&lt;223&gt; a, c, g or t

&lt;400&gt; 106

```

tttccttgaa attaagttca ggtttgtctt tgtgtgtacc aattaatgac aagagggttan 60
atagaagtat gctagatggc aaagagaaat atgttttgtg tcttcaattt tgctaaaaat 120
aaccacagaac atggataatt catttattaa ttgattttgg taagccaagt cctattttgga 180
gaaaattaat agtttttcta aaaaagaatt ttctcaatat cacttggtt gataacattt 240
ttctccttcg agttcctttt tctggagttt aacaaacttg ttctttacaa atagattata 300
ttgactacca ctactgatg ttatgatatt agtttctatt gcttactttg tatttctaatt 360
tttaggattc acaatttagc tggagaacta ttttttaacc tgttgacact aaacatgatt 420
gagctagaag acagttttac catatgcatg ctttttctct gagttatatt ttaaaatcta 480
tacattttct ctaaatatgg aggaatcac tggcatcaaa tgccagtctc agacggaaga 540
cctaaagccc atttctggcc tggagctact tggctttgtg acctatggtg aggcataagt 600
gctctgagtt tgtgttgcc tttttgtaaa atgaggggtt gacttaatca gtgattttca 660
tagcttaaaa ttttttgaa gaacagaact ttttttaaaa acagtttagt gcaaccatat 720
tatataaaac agaacagata caagtagagc taacttgcta aagaaaggat ggaggctctg 780
aagctgtgac ttcatatcc ctttaactct ctatgtcctc tgtagtacct tagatttcta 840
tgggacatcg tttaaaaact attgtttatg cgagagcctt gctaatttcc taaaaattgt 900
ggatacattt tttctcccat gtataattt ctacacttct atttaaaaag aaaaaaaag 960
tcagtgtagt atttaccat tttaccctat aaggagctaa cataactttt gatttagtgt 1020
tattcataaa attaggttag cagtttatta accttttgta tttgctctgg caatgtttta 1080
tatctcataa gctatacaca cctcgaagcc atcaatgaca accttttctt gctgaataga 1140
acagtgattg atgtcatgaa gacaatttta tctccttttg ccttccataa tttgtaccag 1200
gttatataat agtataacac tgccaaggag cggattatct catcttcac cgtgaattcc 1260
agtgtttgtc acgtggttgt tgaataaatg aataaagaat gagaaaacca gaagctctga 1320
tacataatca taatgataat tatttcaatg cacaactacg ggtggtgctg aactagaatc 1380
tatattttct gaaactggct cctctaggat ctactaatga tttaaacta aaagatgaag 1440
ttagtaaaag atcagaaaaa aaagtgggta ttctacaag tcaggacatt ctacgtgact 1500
acaatataat ctacagaaa tttaacatta atacattcta agatttaatt cttagattct 1560
tggtaaacaa attgctcctg tggagatgat tggcatcaca tgggtttttg agctgataca 1620
cccaacactt gagctcactg caaca 1645

```

&lt;210&gt; 107

&lt;211&gt; 2241

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 107

```

gggcctgatt tcagtttctt ccagcccttc ctattgttaa catgggggtt gtgttgaaga 60
atataaagtt acaaagtcaa ggaagtagga aacattttta caagtattat gtagccatct 120
tgggtggagcg gtggtgaggt aggctgcaaa tgattctcct atttctttcc ctgagttcag 180
aacataggaa ttagattgat agacatcaac ataccgctt tattgctgac tcatgacaac 240
taatgggaag acatggctca gatgtgcagc cacagtgagc ttctgaacat ttcttctcag 300
actaagctct tacacacagt tgcagttgaa agaaagaatt gcttgacatg gccacaggag 360
caggcagctt cctgcagaca tgacagtcaa cgcaaactca tgtcactgtg ggcagacaca 420
tgtttgcaaa gagactcaga gccaaacaag cacactcaat gtgctttgcc caaatttacc 480
cattaggtta atcttccctc ctccaagaa gaaagtggag agagcatgag tcttcacatg 540
gaaacttgaa gtcagggaaa tgaaggctca ccaattattt gtgcatgggt ccaagttttc 600
cttgaaatta agttcaggtt tgtctttgtg tgtaccaatt aatgacaaga ggtagatag 660
aagtatgcta gatggcaaag agaaatatgt tttgtgtctt caattttgct aaaaaatacc 720
cagaacatgg ataattcatt tattaattga ttttggttaag ccaagtccta tttggagaaa 780
attaatagtt tttctaaaaa agaattttct caatatcacc tggcttgata acatttttct 840
ccttcogatt cctttttctg gagtttaaca aacttgttct ttacaaatag attatattga 900
ctaccactca ctgatgttat gatattagtt tctattgctt actttgtatt tctaatttta 960
ggattcacia tttagctgga gaactatttt ttaacctgtt gcacctaaac atgattgagc 1020
tagaagacag ttttaccata tgcatgcatt ttctctgagt tatattttta aatctatata 1080
tttctcctaa atatggagga aatcactggc atcaaatgcc agtctcagac ggaagaccta 1140
aagcccatth ctggcctgga gctacttggc tttgtgacct atggtgaggc ataagtgtc 1200

```

```

tgagtttgtg ttgcctcttt tgtaaaatga gggtttgact taatcagtga ttttcatagc 1260
ttaaaatttt ttgaagaac agaacttttt ttaaaaacag ttagatgcaa ccatattata 1320
taaaacagaa cagatacaag tagagctaac ttgctaaaga aaggatggag gctctgaagc 1380
tgtgacttca ttatccctta atactgctat gtccctctgta gtaccttaga tttctatggg 1440
acatcgttta aaaactattg tttatgcgag agccttgcta atttcctaaa aattgtggat 1500
acattttttc tcccatgtat aattttctca ccttctattt aaaaaaaaaa aaaaagtcag 1560
tgtagtattt acatatttta ccctataagg agctaacata acttttgatt tagtgttatt 1620
cataaaatta ggtagcagt ttattaacct tttgtatttg ctctggcaat gtttaataac 1680
tcataagcta tacacacctc gaagccatca atgacaacct tttcttgctg aatagaacag 1740
tgattgatgt catgaagaca attttatctc cttttgcctt ccataatttg taccagggtta 1800
tataatagta taacactgcc aaggagcgga ttatctcatc ttcatcctgt aattccagtg 1860
tttgtcacgt ggttggtgaa taaatgaata aagaatgaga aaaccagaag ctctgatata 1920
taatcataat gataattatt tcaatgcaca actacgggtg gtgctgaact agaattctata 1980
ttttctgaaa ctggctcctc taggatctac taatgattta aatctaaaag atgaagttag 2040
taaagcatca gaaaaaaaag tgggtatttc tacaagtcag gacattctac gtgactacaa 2100
tataatctca cagaaattta acattaatat attctaagat ttaattctta gattcttggg 2160
aaacaaattg ctctgtgga gatgattggc atcacatggt gttttgagct gatacaccca 2220
acacttgagc tcaactgcaac a                                     2241

```

&lt;210&gt; 108

&lt;211&gt; 437

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 108

```

gaataaatga ataaagaatg agaaaaccag aagctctgat acataatcat aatgataatt 60
atttcaatgc acaactacgg gtggtgctga actagaatct atattttctg aaactggctc 120
ctctaggatc tactaatgat ttaaattctaa aagatgaagt tagtaaagca tcagaaaaaa 180
aaggtaaaca aattgctcct gtggagatga ttggcatcac atgggtgtttt gagctgatac 240
acccaacact tgagctcact gcaacagtac cagattttca ccgctatgcc tcctttcact 300
ctgggagctc tccagaggtc ttgcaactcg gagagcatgc tcaggtttcc ccagctctac 360
aaaatcacc cagaatgccaa agacttcaac acaagggtta ataagggtga tctcagaatt 420
gtcacctcaa aaaggcc                                     437

```

&lt;210&gt; 109

&lt;211&gt; 2587

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 109

```

gaggtaggct gcaaatgatt ctctattttc tttccctgag ttcagaacat aggaattaga 60
ttgatagaca tcaacatacc cgctttattg ctgactcatg acaactaatg ggaagacatg 120
gtcagatgt gcagccacag tgagcttctg aacatttctt ctgagactaa gctcttacac 180
acagttgcag ttgaaagaaa gaattgcttg acatggccac aggagcaggc agcttctctg 240
agacatgaca gtcaacgcaa actcatgtca ctgtgggcag acacatgttt gcaaagagac 300
tcagagccaa acaagcacac tcaatgtgct ttgccaaaat ttaccatta ggtaaattct 360
ccctcctccc aagaagaaa tggagagagc atgagtcctc acatggaaac ttgaagtcag 420
ggaaatgaag gctcaccaat tatgtgtgca tgggtccaag ttttccttga aattaagtct 480
agggttgtct ttgtgtgtac caattaatga caagagggtta gatagaagta tgctagatgg 540
caaagagaaa tatgttttgt gtcttcaatt ttgctaaaaa taaccagaa catggataat 600
tcatttatta attgattttg gtaagccaag tcctatttgg agaaaattaa tagtttttct 660
aaaaaagaat tttctcaata tcacctggct tgataacatt tttctccttc gagttccttt 720
ttctggagtt taacaaactt gttctttaca aatagattat attgactacc actcactgat 780
gttatgatat tagtttctat tgcttacttt gtatttctaa ttttaggatt cacaatttag 840
ctggagaact attttttaac ctggttgacc taaacatgat tgagctagaa gacagtttta 900
ccatatgcat gcattttctc tgagttatat tttaaaatct atacatttct cctaaatatg 960
gaggaaatca ctggcatcaa atgccagtc cagacggaag acctaaagcc catttctggc 1020
ctggagctac ttggctttgt gacctatggt gaggcataag tgctctgagt ttgtgtggc 1080
tcttttgtaa aatgagggtt tgacttaatc agtgattttc atagcttaaa atttttttga 1140

```



```

agaacagaac tttttttaaa aacagttaga tgcaaccata ttatataaaa cagaacagat 1200
acaagtagag ctaacttgct aaagaaagga tggaggtctt gaagctgtga cttcattatc 1260
ccttaatact gctatgtcct ctgtagtacc ttagatttct atgggacatc gtttaaaaac 1320
tattgtttat gcgagagcct tgctaatttc ctaaaaattg tggatacatt ttttctccca 1380
tgtataattt tctcaccttc tatttaaaaa aaaaaaaaaa gtcagtgtag tatttacata 1440
ttttacccta taaggagcta acataacttt tgatttagtg ttattcataa aattagggtta 1500
gcagtttatt aaccttttgt atttgctctg gcaatgttta atatctcata agctatacac 1560
acctcgaagc catcaatgac aaccttttct tgctgaatag aacagtgatt gatgtcatga 1620
agacaatttt atctcctttt gccttccata atttgtacca ggttatataa tagtataaca 1680
ctgccaaagga gcggtatttc tcatcttcat cctgtaattc cagtgtttgt cacgtggttg 1740
ttgaataaat gaataaagaa tgagaaaacc agaagctctg atacataatc ataatagataa 1800
ttatttcaat gcacaactac ggggtgggtgct gaactagaat ctatatatttc tgaaactggc 1860
tcctctagga tctactaatg atttaaatct aaaagatgaa gttagtaaag catcagaaaa 1920
aaaaggtaaa caaattgctc ctgtggagat gattgggcat cacatggtgt tttgagctga 1980
tacaccaac acttgagctc actggcaaca gtaccagatt ttcaccgcta tgcctccttt 2040
cactctggga gtcttccaga ggtcttgac tcgggagagc atgctcaggt ttccccagct 2100
ctacaaaatc acccagaatg ccaaagactt caacacaagg gtaaataagg ttgatctcag 2160
aattgtcacc tcaaaaaggc cctgccttcc actgttcagt tctgggtcatc tgcctatgag 2220
atatctgaag cttgaaagag aacacttgaa aatcactgag accgtgactc ccatcccagc 2280
acacacagca agccaagtag gttacagaga tttcttcttg ggtgatgagt tcacgccaca 2340
tggggtatgt tctccagttc cagtgtgttg actccttctt gcttcccccc atcagccctt 2400
gaggtcaatg tgggcagcag ccccatggt ccaagttcta gatccgctgt ggaagacttt 2460
tttaggcaat cacacagccc tgcattggag cctgatgaga gcttgccata ttgttgctag 2520
gtttgtcatt ttaaatacag tgtttcttta gctagttagt aaaattggct atataggaaa 2580
aaaaaaa

```

```

<210> 110
<211> 448
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (409)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (431)
<223> a, c, g or t

```

```

<400> 110
agactacctg gtatggccta agggccccc aa gtaaattaaa acattcatat caggcaggat 60
agtccaagga cttagagggt atcttgaatc tggacacagg acaaacattt ctttggaaatg 120
tgaaaagttt acacaatcaa gacttgctat gttgatcctt tactgcggta ttgttttttg 180
ttttctgacc caaaagtgat gtgcattgct tacctatcat ggccttggtt aagggtgtcag 240
tcttgccctc aaatttcctc ctaaagcatg gaattggcca taaatttgct tagtgatttt 300
ctgatttagt atcattagtt tgatgactag ttttattatg tgagtgtgat aaaagggttac 360
gcttagtata caaagacctt cttataagtc tagttaataa aactagggnc atgtgttact 420
acaacatagg ngttaacttt gtctggag
448

```

```

<210> 111
<211> 798
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (770)

```

<223> a, c, g or t

<220>

<221> unsure

<222> (773)

<223> a, c, g or t

<220>

<221> unsure

<222> (779)

<223> a, c, g or t

<400> 111

```

agactacctg gtatggccta agggccccc aa gtaaattaaa acattcatat caggcaggat 60
agtccaagga cttagagggt atcttgaatc tggacacagg acaaacattt ctttgggaatg 120
tgaaaaagttt acacaatcaa gacttgctat gttgatcctt tactgcggta ttgttttttg 180
ttttctgacc caaaagtgat gtgcattgct tacctatcat ggccttgggt aagggtgtcag 240
tcttgctccc aaatttcctc ctaaagcatg gaattggcca taaatttgct tagtgatttt 300
ctgatttagt atcattagtt tgatgactag ttttattatg tgagtgtgat aaaagggttac 360
gcttagtata caaagacctt cttataagtc tagttaataa aactagggcc atgtgttact 420
acaacatagg agttaacttt gtctggaggc tttttcaagc ccaagagggt gtcatctctt 480
tatgtgtaag atactgcca catactatac tgaagtcagg aaccaagaac cggtaatttt 540
acctcagcca agttgcttgg actgctttgc ttaataatgg gtttgccata tctactttaa 600
tgagatgtgt ggcattggta tggttacaaa agtgatttgt taattattgc ctggtaggaa 660
ggagaacatg ttttttttga acctatgcaa atagccacat gtctgtgaaa agtaaaggga 720
tacttttgtg taaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaattn ctnggtcgnc 780
aaggcaattc gtggtcgt 798

```

<210> 112

<211> 683

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (676)

<223> a, c, g or t

<400> 112

```

gggaaactgg gttccggagg gtcagagctg cctctgcttc aggcctcttc atgatctgga 60
tcaagcttgt ttgacttcca tctacagaag tcaaccttgg cttctcaaag agcaaaatag 120
ggctgagcag actagccctg gggaagggtc ctgtgtccta aggcctggggg aaccaaggga 180
agaggtttgt gttatctgga tttggaagct ggaagaaggg accctacagg gctgagactc 240
aggcttctga ggagggggaa ctgcccagct aaaactgggt cctttgccag ggtacaatga 300
ggctaattct gggaacatag aaagaagctg acctggaacc agctacagcc accagtgtcc 360
aatacacagt tagcagggtc tcttagagag aaagactcaa aagattgtct gtgaataccc 420
taaataccaat ccagggtatac tggatgcaaa cactaggaga aaaaagaaag cttttatata 480
aggaattaaa ttgccctctg tctgaagaaa gaaacgatta ggaaagatgg aagtgtctctg 540
aataaaattt acaaaaggag aatcagacag accaacctgg gtgatgtgtt tcaagctctg 600
ccccaggagc ttgaatgtct gttccttgtg gcaggagcat tcctagagac tgggtgcctca 660
ggttttatgt tggtnctctc ctg 683

```

<210> 113

<211> 735

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (698)  
<223> a, c, g or t

<220>  
<221> unsure  
<222> (700)  
<223> a, c, g or t

<220>  
<221> unsure  
<222> (704)  
<223> a, c, g or t

<220>  
<221> unsure  
<222> (709)  
<223> a, c, g or t

<220>  
<221> unsure  
<222> (714)  
<223> a, c, g or t

<220>  
<221> unsure  
<222> (730)  
<223> a, c, g or t

<400> 113  
gggaaactgg gttccggagg gtcagagctg cctctgcttc aggccttctc atgatctgga 60  
tcaagcttgt ttgacttcca tctacagaag tcaaccttgg ctctctcaaag agcaaaatag 120  
ggctgagcag actagccctg gggaaggtca ctgtgtccta aggcctggggg aaccaaggga 180  
agagggttgt gttatctgga tttggaagct ggaagaaggg accctacagg gctgagactc 240  
aggccttctga ggagggggaa ctgcccagct aaaactgggt cctttgccag ggtacaatga 300  
ggctaattct gggaacatag aaagaagctg acctggaacc agctacagcc accagtgtcc 360  
aatacacagt tagcaggtca tcttagagag aaagactcaa agattgtct gtgaataccc 420  
taaatccaat ccagggtatac tggatgcaaa cactaggaga aaaaagaaag cttttatata 480  
aggaattaaa ttgccctctg tctgaagaaa gaaacgatta ggaaagatgg aagtgatctg 540  
aataaaattt acaaaaggag aattcagaca gaccaacctg ggtgatgtgt ttcaagctct 600  
gccccaggag cttgaatgtc tgttccttgt ggcaggagca ttcctagaga ctggtgcctc 660  
aggttttatg ttgttgtctc tctgctccag ctcaacctn aaanctctnc cctnggatgt 720  
tggaactcan attct 735

<210> 114  
<211> 601  
<212> DNA  
<213> Homo sapiens

<400> 114  
attttgagga atttacttgg atctttcaag atcctgctgt agctaagcaa gatcttcaag 60  
aaaacccttg tcttgccctc atttgttaaa tgcaatttgc cttgtcaaag gactaggagg 120  
ccagtatagc aaggtccctt tgggaaactg ggtccggag ggtcagagct gcctctgctt 180  
caggcttctt catgatctgg atcaagcttg tttgacttcc atctacagaa gtcaaccttg 240  
gcttctcaaa gagcaaaata gggctgagca gactagccct ggggaaggct actgtgtcct 300  
aaggctgggg gaaccaaggg aagaggttgg tgttatctgg atttgaggc tggagaagg 360  
gacctgcag ggctgagact caggcttctg aggaggggga actgcccagc tagggctgat 420  
acctttgccg ggggtcaatg aggctaattc tgggaatata gaaggaagct gacctggaac 480  
cagctacagc caccagtgtc caatacacag ttagcaggtc atcttagaga gaaagactca 540  
aaagattgtc tgtgaatacc ctaaatccaa tccagggtata ctggatgcaa acactaggag 600  
t 601

<210> 115  
 <211> 334  
 <212> DNA  
 <213> Homo sapiens

<400> 115  
 gtttgaagg gctcaactca tgccaatata cccttgccct tgcagagtgg gacgcaagaa 60  
 cctgatgtta gcaaaactctc cacacttta ttcaaccctg caaacacttt ccaagtgcct 120  
 tctgtttgtc aggcaatatg ccagtcacta ggaatgaaga ggtagataag gatgggcctt 180  
 aaaatcaatc tagtgaggag ggacaacata aacaaatgaa caaacaatta tcagaaatat 240  
 tttttatggg gtggcaaaac agtgaaatat aggatctaatt ttggtttgga aagaaagaga 300  
 aaatttaaca gaaaaagtaa tgtttgaagc cagt 334

<210> 116  
 <211> 193  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (38)..(90)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (94)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (96)..(97)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (99)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (156)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (172)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (179)  
 <223> a, c, g or t

<400> 116  
 gaatgaaaca gtgggatgca gtgaggaaga ggaacacannnn nnnnnnnnnnn 60  
 nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn tagngnncna atgtaggggg aaagttaata 120  
 aaaattgtat tatattaggc gttttttgtt aaatangtag atcatagctg cncttgtcnc 180  
 aaaaaggtaa cta 193

<210> 117  
 <211> 152  
 <212> DNA  
 <213> Homo sapiens

<400> 117  
 ggcatagccct atgattctac gtatattaag tccaaacatc aagcagtact atctatgata 60  
 gtaaaactgg taggctaaac attggggagt tattacattg acaacaacat gaaggggttt 120  
 atgggacttc tggtaatgtt ttgtttcctg at 152

<210> 118  
 <211> 498  
 <212> DNA  
 <213> Homo sapiens

<400> 118  
 atatatTTTT gaatgaagtt caaggcatgt atTTTTatgt gttaaccaa agatgcattc 60  
 aaggatgtta atctcagcag gatttgtaat acaccaaatt tggaaaaaac cctcaaaatg 120  
 tctatcaatg gtagaatgga atTTTataca gcaatggaaa tgcattgaact atgactatta 180  
 gcagcaacat gaatgatttt cataaaaaata gTTTTgagca aagaatccag atataaaaga 240  
 gggcatgccc tatgattcta cgtatattaa gtccaaacac aagcagtact atctatgata 300  
 gtaaaatggg aggctaaaca ttggggaggt attacattga caacaacatg aaggggttta 360  
 tgggattctg gtaatgtttt gtttcctgat ctgggcactg ggtaccagga tgtatttact 420  
 ttgtgaacat ttaaccatct acgatttttg gatttttctg ggtatatgtt atactttcaa 480  
 taaaacattt tcatgaga 498

<210> 119  
 <211> 663  
 <212> DNA  
 <213> Homo sapiens

<400> 119  
 atgtatttga ctatattatt tagttctgaa aggaattttt taatacagtg aaacattagt 60  
 taaaaataaa tgttaactcc aaaaagaata ataaaatcat tattttttta tcagaggctg 120  
 ccagtgtttt tctgcttaat ctatacagca tcttagctgt aaaaatgata gaattcatgt 180  
 tattttacta tgcattctca gtaacaataa atgtgttgct cacttctaca ttttagatag 240  
 ctgacataag gcctaacatg gaaagcaaaa agctatccag gaaatatttc cctgctcatg 300  
 catatttcct ttggcataca aatcattgta aatgatgggt agttaacttc aaacatttca 360  
 agctacacaa ctaatgtaat taaaccctaa gtccaattta ctattggccc ttggtacttc 420  
 tgcaggctga tgttaagtgt catttttcaa gtctttcatt gcctataaac aagatgggtg 480  
 caattttctt tattttctga acacttgcaa ttctctaaaa gtttcatact tcttcatacc 540  
 ttcaaataatg atttaggctt ttctagtacg taggcttccc ttttctgcct tgcttgaga 600  
 ccctgttcat gcttggttaga catagcttca acgtgacatc tgggaagctt cccatgacat 660  
 gga 663

<210> 120  
 <211> 904  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (684)  
 <223> a, c, g or t  
 <220>

<221> unsure  
 <222> (888)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (893)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (896)  
 <223> a, c, g or t

<400> 120  
 aaaaaaaagg actttgagtc cattcaaagt taagtaggag ctctccaggt tcttccagtg 60  
 acccatttac cacctctact cctcacctca catctggctt cctccagggg ccctgataca 120  
 gtgggtgatg ggtcctaagg gggcctccag gaccaccag ccctatgagg aaagagttct 180  
 tcctgatcct accccttgac ttcttttct ttctcctgca ggtctcagaa cggccccgaa 240  
 gcctccccct gtccctgaa ttggagagct ctcttgatg ccctctgtta gggccccacc 300  
 caatcccagg gcagaaggac atgaggagc aaagagcttg aggaatgcca tactccggct 360  
 ggtccgggac atggaaattc ggactcaggg aggacccggg ctgggcaatg actgggagac 420  
 ttgcctgggt tcccaggact tgggggtcct gactcccagc cctcatcctg cagtcccctc 480  
 tgttcccagc cccagccttt ctaagccatt gggaaatagaa tggccccctt tgttctgggtg 540  
 tccaggggtg atttgtgcaa agctcttatt tccagtggca agccccaga ggcttgtaag 600  
 agttgggatg agggatggag agggactggg tctctgggaa cagggttgag gtcttatctg 660  
 tggactgtct gactcccagc tggagccaag atggggcatg tccccgtctc tgcttagcgt 720  
 ctgggtgaga aaaacaggct gtgatccaga agaagggaag atagagaagg agggaaagga 780  
 tgtaggcgaa ggaggtgaga gacaggatag gaggaaggaa gtggaggagg aggtggtagg 840  
 aattggaatg aggtagaagc cgtgcagagg aagaggggag aaggacgnag gangancgat 900  
 gaag 904

<210> 121  
 <211> 1309  
 <212> DNA  
 <213> Homo sapiens

<400> 121  
 aaaaaaaagg actttgagtc cattcaaagt taagtaggag ctctccaggt tcttccagtg 60  
 acccatttac cacctctact cctcacctca catctggctt cctccagggg ccctgataca 120  
 gtgggtgatg ggtcctaagg gggcctccag gaccaccag ccctatgagg aaagagttct 180  
 tcctgatcct accccttgac ttcttttct ttctcctgca ggtctcagaa cggccccgaa 240  
 gcctccccct atccctgaa ttggagagct ctcttgatg ccctctgtta gggccccacc 300  
 caatcccagg gcagaaggac atgaggagc aaagagcttg aggaatgcca tactccggct 360  
 ggtccgggac atggaaattc ggactcaggg aggacccggg ctgggcaatg actgggagac 420  
 ttgcctgggt tcccaggact tgggggtcct gactcccagc cctcatcctg cagtcccctc 480  
 tgttcccagc cccagccttt ctaagccatt gggaaatagaa tggccccctt tgttctgggtg 540  
 tccaggggtg atttgtgcaa agctcttatt tccagtggca agccccaga ggcttgtaag 600  
 agttgggatg agggatggag agggactggg tctctgggaa cagggttgag gtcttatctg 660  
 tggactgtct gactcccagc tggagccaag atggggcatg tccccgtctc tgcttagcgt 720  
 ctgggtgaga aaaacaggct gtgatccaga agaagggaag atagagaagg agggaaagga 780  
 tgtaggcgaa ggaggtgaga gacaggatag gaggaaggaa gtggaggagg aggtggtagg 840  
 aattggaagg aggtagaagc cgtgcagagg aagaggggag agggacgaag gaggagcgat 900  
 gaagaagagg agggagacaa aaagagggat ggaggagaga gggagtctgg agaacaagg 960  
 gtcttttctc tggggagggg tgcagtgggc ggggctgaca ctgtcagcca atctcccat 1020  
 cggggaagag aatcctggac agggacagga tggggagggg atttataacg ggcttttttg 1080  
 tgggagatgg gtacccagtg ggggcccactg gaggtctcc gggcacactc tggcccttcc 1140  
 cagaaaaggg gtccgtcttc tcgaatcctt ccacagttgt gtattgcaaa ctacggcgca 1200  
 ttttactatt gatcacacgt cattatcttg tcattacata ctatttctat tcaacctccc 1260  
 ccaactgaag tgtggccgcc acaatcacca ccaaccccca cacaaccaa 1309

<210> 122  
 <211> 295  
 <212> DNA  
 <213> Homo sapiens

<400> 122  
 gcagtttggg aattagaaaa aaacccaaag tatgcatgca gttctgtaag ataaagtgtc 60  
 tgtccaggca tgcatacaac ccagcaattg catgcctggg cgcttacctt acagaaatga 120  
 acatttataa ttacattata atatgtacac caaattcatc acagctttat taatagaagc 180  
 caaactctct gtgggcttct cacagtgtac ccattgccag agtaaaactgc agccttgaac 240  
 cattgctcag cctccttacc catgagctat gaacactgaa gcaggttgca cagtg 295

<210> 123  
 <211> 714  
 <212> DNA  
 <213> Homo sapiens

<400> 123  
 caaaagaccc tgtcaacagg attaaatata cagccaccga ctccggagaaa atattttgca 60  
 aaacagcgta tccggcaaag aattaatatc agaatacata atgaattctc gaaaactcga 120  
 agtataacaa atatcgtaca ctggaaagtg ggcaaaacca ttaataaaca tttcaccaca 180  
 taggatatat agatggccaa agaagcatat gaaaaagatg cgcaacatca ttagctatta 240  
 gggaatgca aattaaaacc accattagga tattagtaca gaatggttaa acatcaaaaa 300  
 taatagtgat aacaccaaatt gccataaagg aagtggagga gaaataggat cattgatata 360  
 ttgttttttg gaaggtaaaa tggtagacc ctctagaaag cagtttggtt attagaaaaa 420  
 aacccaaagt atgcatgcag ttctgtaaga taaagtgtct gtccaggcat gcatacaacc 480  
 cagcaattgc atgcctgggc gcttacctta cagaaatgaa catttataat tacattataa 540  
 tatgtacaca aaattcatca cagctttatt aatagaagcc aaactctctg tgggcttctc 600  
 acagtgtacc cattgccaga gtaaaactgca gccttgaacc attgctcagc ctcttacc 660  
 atgagctatg aacactgaag caggttgac agtgaaaaaa aaaaaaagtc gacc 714

<210> 124  
 <211> 924  
 <212> DNA  
 <213> Homo sapiens

<400> 124  
 agagtggcct aggacagctc ctctcctgcc agagctaggc aggcgccgaa gtagccgcat 60  
 ggccccgtca gaagacccca gggactggag agccaacctc aaaggcacca tccgtgagac 120  
 aggcctggag accagctccg gtgggaagct ggctggccat cagaagaccg tccccacggc 180  
 tcacctgact tttgttattg actgcaccca cgggaagcag ctctccctgg cagcaaccgc 240  
 atcaccaccc caagccccca gtcccaatcg agggcttgtc accccaccaa tgaagacctt 300  
 catcgtgttc tgtggggaaa actggcccca tctgactcgg gtgaccccca tgggtggggg 360  
 atgccttgcc caggccaggg ccacctgcc gctctgcaga gggtctgtgg cctcagcttc 420  
 ctccccagtc agcccgtctc gccccagga ggttcccgag gctaagggga aaccctgtaa 480  
 ggctgcgcct gtgaggtctt caacttgggg aacagtcaag gactcactga aagccctctc 540  
 ctcttgtgtc tgtgggcagg ccgattagct ggaaggaccg ggctctgatg ccagaggct 600  
 gcaattccca gggcctggcc ctgcttcccc agctaagcag gagtcttttg tgcttgagcc 660  
 aaggaaacat cattagatcc gctaaggggc atctgaacaa tccgtcgagt ggcagaggca 720  
 ggataagtca cctgcacatg aagagactca ttcattcata cagcaaatat tactggtaga 780  
 tcttccacat gccaggccct gcaaagtgtt ggggagatac catggtttac atggagctgg 840  
 tatttttggg gtggagggaa cccaccctga ataaataaag taaccaata aataaagaag 900  
 atgattttga acagcgaaaa aaaa 924

<210> 125  
 <211> 939

<212> DNA  
<213> Homo sapiens

<400> 125  
agagtggcct aggcagctc ctctcctgcc agagctaggg aggcgccgaa gtagccgcat 60  
ggccccgtca gaagacccca gggactggag agccaacctc aaaggcacca tccgtgagac 120  
aggcctggag accagctccg gtgggaagct ggctggccat cagaagaccg tccccacggc 180  
tcacctgact ttgtttattg actgcaccca cgggaagcag ctctccctgg cagcaaccgc 240  
atcaccaccc caagcccccga gtcccaatcg agggcttgct accccaccaa tgaagaccta 300  
catcgtgttc tgtggggaaa actggcccca tctgactcgg gtgaccccca tgggtggggg 360  
atgccttgcc caggccaggg ccacctgcc gctctgcaga gggctctgtg cctcagcttc 420  
cttcccagtc agcccgctct gccccagga ggttcccag gctaagggga aaccctgtga 480  
ggctgcgcct gtgaggtctt caacttgggg aacagtcaag gactcactga aagccctctc 540  
ctcttggtgc tgtgggcagg ccgattagct ggaaggggcg ggctctgatg cccagaggct 600  
gcaattccca gggcctggcc ctgcttcccc agctaagcag gagtcttttg tgcctgagcc 660  
aaggaaacat cattagatcc gctaaggggc atctgaaaca tccgtcgagt ggcagaggca 720  
ggataagtca cctgcacatg aagagactca ttcattcata cagcaaatat tactgggtaca 780  
tcttccacat gccaggccct gcaaagtgtc ggggagatac catgggtttc ctggagctgg 840  
tatttttggg gtggagggaa cccacctga ataaataaag taaccaata aataaagaag 900  
atgattttga acagcgaaaa aaaaattcga gctcgttg 939

<210> 126  
<211> 317  
<212> DNA  
<213> Homo sapiens

<400> 126  
aaaaaggttt gaatatttaa aatcagttcc atgttagtca aagagttaca attatagttc 60  
aactaaacct gcagtcaatg taagtattca taccctaaga aaaagcacca caaatgatg 120  
tctgtgattg ttaacggttg attggtttcc tgtgtccata gtggacaata ttatgaagca 180  
tagacagaaa aacatgttta ctaagaagct ttttttccct tccaggaaat tctgtagggtg 240  
aaacatgttg aacattgtca gttgacacat attctggtga agtctaacat taaacattaa 300  
actaaaaagc aagtgac 317

<210> 127  
<211> 144  
<212> DNA  
<213> Homo sapiens

<400> 127  
gaaaccaa attttccctaa tgcaaaagtt gaaaccttat ctgaatggaa gttcgtagt 60  
taattcccca agttttaaat gcctgggctt cgtaatatgag tttttaccaa ttgtctgcta 120  
catgtgtaaa atttcacctc agca 144

<210> 128  
<211> 161  
<212> DNA  
<213> Homo sapiens

<400> 128  
aaccaaatat tttcctaagt caaaagttga aaccttatct gaatggaagt tcgtagtgt 60  
attccccaag ttttaaagtc ctgggcttcg ttaatgagtt tttaccaatt gtctgtctaca 120  
tgtgtaaaaa ttcacctcag cattttgtgg ttttgtttt t 161

<210> 129  
<211> 728  
<212> DNA



<213> Homo sapiens

<220>

<221> unsure

<222> (239)

<223> a, c, g or t

<220>

<221> unsure

<222> (255)

<223> a, c, g or t

<400> 129

```
gcagagtttag agccaatttc atttgcctcg cactgagaat gtgcattccc agcagatcat 60
ggttgagctc agcacacctg caactcagcc agcctctttg aagggtgcac agttactaac 120
tgtcagcggt gcacagccac ctttagcagg tatgttcaga acttacattc ccaccctttc 180
taaaacagct cacagtaaca agagaacggg atttaacttt tgacatgcac actcatgana 240
accaatgatt tttgnaagcc aaattgttga gaagataagg tggggattct gactactagt 300
atttttacaa atctgattgt cgttgcagggt tttgttttat tttgttgtgt taatgctgag 360
agtggagaat agattggaat atttgcctct tgtgtttctt tttgctttgt aacattgcaa 420
gtgggtccaca ttttccttta tttaaaattt aaagtgtgtg cctgggggtt ctggtgttaa 480
atagaagtga tacttctgca taaagtatta tggagatgct ctgttccatc caggagggtg 540
cagggtgaaa gagggcagtt catcctctc acacttaggg caggagcag catgcagggg 600
cagatcagca gctttgcatg ttgacaaatc cactctctgc tgcagatgcc taggggaagt 660
tgcagactta aattttcttt tgtaaaatgg gggaacacaa acagatctta tgtcactggt 720
tactccag 728
```

<210> 130

<211> 680

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (375)..(563)

<223> a, c, g or t

<220>

<221> unsure

<222> (615)

<223> a, c, g or t

<220>

<221> unsure

<222> (649)

<223> a, c, g or t

<400> 130

```
ttccgaccta cattataaga ctgctgtaat gatctaaaac tttagctggt ttaatatagt 60
tttaaaacta atgatatctt tctctgtcag taaaatacaa actttttctt aataaaaaatg 120
taatggaaaa ctgttcctca tagatttttg tcactttaca aagtgacaaa atcattttgt 180
tagtttatgg aaaataagct tgtaaaactt ttacctaata agataggact gaaatttcag 240
cttttttaat ttgatgatga gtttttaatt tcttttgaaa agaattgtat gcttctaata 300
atztatcaag aggaagaata ccaaaagaaa atatctgctc ttctttcttt ttacttagat 360
ttttttgcat ttttntntnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
nnnnnnnnnn nnnnnnnnnn nnnactgttt ttctctgct ctcatacaac aatcaacaca 600
gaaggcttct gtggnctcaa atgtggggga attttttcca cgccaagcna gcagtcaggg 660
ctgcagtgtc tccaacttag 680
```

<210> 131  
 <211> 858  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (375)..(563)  
 <223> a, c, g or t

<400> 131  
 ttccgaccta cattataaga ctgctgtaat gatctaaaac tttagctggt ttaatatagt 60  
 tttaaaacta atgatatctt tctctgtcag taaaatacaa actttttctt aataaaaatg 120  
 taatggaaaa ctgttcctca tagatttttg tcactttaca aagtgacaaa atcattttgt 180  
 tagtttatgg aaaataagct tgtaaaactt tttacctaaa agataggact gaaatttcag 240  
 cttttttaat ttgatgatga gtttttaatt tcttttgaaa aagaatgtat gcttctaata 300  
 atttatcaag aggaagaata ccaaaagaaa atatctgctc ttctttcttt ttacttagat 360  
 ttttttgcatt ttttnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420  
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480  
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540  
 nnnnnnnnnn nnnnnnnnnn nnnactgttt ttctctgctc ctcatataac aatcaacaca 600  
 gaagacttct gtgaccaagt ggagtccttt tatgacacat tccatactgt ggctgacatg 660  
 atgtatttct gccagatgct ggcagttgtg gaaactatca atgcagcaat tggagtcact 720  
 acgtcacccg tgctgccttc tctgatccag gtctcacatc caggtcacgc tgatgcaaaa 780  
 ggttttatata taaattttcg tcacctttat aaacagcgca gacggcgcta tggacaaaaa 840  
 aagaaaaaga tccactaa 858

<210> 132  
 <211> 328  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (107)..(240)  
 <223> a, c, g or t

<220>  
 <221> unsure  
 <222> (254)..(255)  
 <223> a, c, g or t

<400> 132  
 aatttgtatg aactattttt taagatatta actcttttct cctatatattt tgttttcatt 60  
 taaacattgt tgtgacatat tgaaatgttt atctttactt ttcattnnnn nnnnnnnnnn 120  
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180  
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240  
 tgcttttttc cttnttggtt tcttctattc cttttgagat cagtttcctt ctgttaaggg 300  
 ctgaaatgtg gtccccctca aaaactcc 328

<210> 133  
 <211> 762  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure

&lt;222&gt; (593) .. (706)

&lt;223&gt; a, c, g or t

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (748)

&lt;223&gt; a, c, g or t

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (751)

&lt;223&gt; a, c, g or t

&lt;400&gt; 133

```

cctacacact ggtaaatgcaa cagaatgcc aagagtgacc tcataaagca aggattccct 60
tcgtggcccc ttctctgctg cctctcagaa tccagacgct aaggaaaatc cctaagcaga 120
gattttctgt tggatgctaa aagcaaggaa taaaagttga aaatttgga aatgtctcaa 180
caccgtcacc agcgccactc gagagtcatt tctagtccac cagttgacac tacatcggtg 240
ggattttgcc caacattcaa gaaatttaag taaatattat ctatctccat tgctgttaa 300
gaaatgtgct agtagaagtg tgagggcagg gtgtcagtgt tctctcagcc tcttccctca 360
gatactcgtc tgcttaccaa aataagttgc atgtccttga caatctggtt tctatgattg 420
gtgaggctgg catgctatta cctttatgtg ccctgtagac ttgaatgacc agtttgacca 480
gtttgactgt tagataatct gaaggctttt ctcttttttt ataatagacc ccatctcaaa 540
tcagataatg aaaattacat atcttgatat attagaaaag tatatacatt ctnnnnnnnn 600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnngaga gtacacagat 720
tggcccgagg tgatggtgtg cgcctgtngt nccggctgct cg 762

```

&lt;210&gt; 134

&lt;211&gt; 11677

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 134

```

cagccctcgg cagacggcca atggcgggcg tgctcggggc gctcggggcg acgcggcgct 60
tggtggcggc gctgcgaggg cagagcctag ggctagcggc catgtcatca ggtactcaca 120
ggttgattgc agaggagagg aaccaagcta tacttgacct taaagcagca ggatggtcgg 180
aattaagtga gagagatgcc atctacaaag aattctcctt ccacaatttt aatcaggcat 240
ttggctttat gtcccagatt gccctacaag cagagaagat gaatcatcac ccagaatggg 300
tcaatgtata caacaaggct cagataactc tcacctcaca tgactgtggt gaactgacca 360
aaaaagatgt gaagctggcc aagtttatgt aaaaagcagc tgcttctgtg tgatttcttc 420
caaaatacat aagtctgaga ggctaaactt gatggctgtg ttaacatatg tcacgtgtag 480
cacagtggag aaagcaggat atggctcata atgacagtgg tgaagacctg cgaatgaagt 540
tgctagttaa cacctacatt agggtttgac ataggtctat gttatgggtc gctgcactcg 600
ctggaactca cagactttac tatagagaat caaagatccc gtatccgaag tctatggaaa 660
tgctcatggt ggtaaattcc aacagaatga aacaccaaac ttgcttaag taactcacgt 720
ttcaatttga aagagatatt gtcaaaattg gaggcccca gggtcctgtc tggtccaaat 780
ctttgcatga tgacagtggg ttctctgatg tggtaagctt tggctttctt ctgttttctt 840
tctaaaagat cactggagta gagaggagt aaacagacat gacctttgac ctcttgcatt 900
acctccacag atagcaaacc gggccgacac atggttgacg atgtcctttt ctacaatgaa 960
gttaatgaaa gttctgaaaa tagtgattac tttctgacat tgataggatt taggaaacct 1020
ctggataaat agcttaagca tggctgttta tgtttttgct atagacaaaa agcagcagca 1080
tgtacattgt atttggacac aagcctgcct cggtaatat attgaactat tggaccacta 1140
gggttagtag ggagcgggtc gtacactttc tgattcagca ttcagaaaca ttctaggtgg 1200
actctgtagc tttcagtttt gtaaagtatt cggaaaaaca tcgggagggt ttggccatca 1260
tatgtgagct ttgtgtttca atgccagtta ctcaggatta gtaaattaat gactgtccag 1320
aggacttcag ggtcaccaag ctgctgcacc tgccattggc tgactctccc cggctatctg 1380
tggctgagat ggtgctgctt aggtcacgca gagcatgagc tgctgtgtaa agggcacagg 1440
agatggccct tgggcttctc atccaggat gcttgccctg cccaccaatc catgagaaga 1500
tatgtatgat ttcagtaggc cctggatcag cttgtcacct ctgggttctt gtttgctttc 1560

```

cactcactca	gctggagttt	catttccaga	ctaaagtctt	catcattggc	ttcagaaaca	1620
gcattcatct	gtggctgtgc	tgatgtagta	caccaagaac	aactgggctc	ttctctgtca	1680
ctttcagtgg	gctaccttcc	ctcacctctc	caagcagcat	gaaagaattc	tttacatttt	1740
taatctcttt	tttgtttttc	octgaaagta	tgctttgggtg	cttaaagaga	gaagtcacaa	1800
aagtatacta	ctgagtttcc	tggagatgaa	atcctgttgt	ccctagctat	gtgaatgagc	1860
acagggatcc	ctgatgccat	tattttgtat	attcatacgg	cacacactta	ctgagggcct	1920
tctgtgtgcc	ctaggggatt	gagcacagtg	acatatcagg	gcaggtagaa	acagatggag	1980
agctgatgcg	ggctgtctta	gagcagctgc	cccaggaggc	ccctgtggat	ggatgttggg	2040
caggagccct	gagacgttag	gggcatataa	ctaaaggaca	tagcaggagt	tataggagga	2100
gctgatccct	gagggaaaaca	atgaagacgg	agaagatggg	gctaaaagttt	gaattgtggg	2160
gacattaatc	acggtgattc	ttaaaacttt	gctgttgatg	attttaaatg	gagaaaaatga	2220
gtacgtaaga	tggtatttcc	cagttcagta	tataggttgc	ccacaaagta	ttttcctacc	2280
atgaatggtc	atatatactt	gttgtagaat	accaggggaca	gcagagatgg	tggggtagt	2340
acttcccttt	cttacagccc	aagaactttg	gtgtccagga	gattgaccaa	tttagccact	2400
gagcattttc	tacaacacag	ggctaccag	atcccactgt	cctgatttgc	cctgaaagcc	2460
aaagggttca	ggagaagggtg	agtggggtga	atatattaat	cctgagagtt	gaacagagca	2520
aaaaatcccta	ttacttttgt	acttaaaaca	tctctgccac	atgtgctcac	tctttatatt	2580
ctgttttaggt	ggtttatatg	tgcacatccc	atcctatgcc	tgtagttagc	caactcaggg	2640
tttatattgc	ctcctttctt	tttttctttt	tttttttttt	ttttaagaga	tggggtctcg	2700
ttctgtcatg	cagactggag	tgtagtgggtg	tgatcacagc	tcattgtaac	ctccaacgcc	2760
tgtagtgaag	tgatcctcct	gccttggcct	ctctggtagc	tgggactaca	ggtgcatgcc	2820
accacaccca	cctaattttt	tttattttta	ttttttgtag	agacagtctc	actatcctgc	2880
tcgggctggg	cctgaactcc	tgggctcaag	ttatcttgct	gcctcagcct	cccatgggta	2940
atctttatct	cctttttttt	ttttttttgg	agatggagtt	tcgctcttgt	cgccagggt	3000
ggagtgcatt	ggcagatct	tggtcactg	cagtctccac	ctcctgggtt	caagtgattc	3060
tccatcctca	gcctcctgag	tagctgagat	tacaggcaac	tgccaccatg	cgcggtctaa	3120
ttatgtattt	tttttttagta	agagatgggg	tttcaccatg	ttggccagac	tggtcttaaa	3180
ctcctgacct	caagcgacct	gcctgccttg	gcctcccaaa	gtgctgggat	tacaggcatg	3240
agccgctatg	cctcgctcgt	gatttttatt	tcttattttt	tttttagaga	tggggtctc	3300
actatgttgc	tcaggctgat	ctcaaaactcc	tggcctcaag	tgatcctccc	accttagcct	3360
cccaagttgc	tgggattata	agtgtgagcc	actatcccta	cctcactatt	accttctttg	3420
cttctcttgt	tttcttttgt	tctaagtcac	acctatcaca	atcttttctt	gtccttccag	3480
gtgttttcca	gtgctgtgcc	ctggatgtgc	tctctttctc	ttagagccca	gagaacttgc	3540
ttttccccct	tatatatgac	ccttaacttt	ttctaacaca	ttattaaggg	cctgtgtcta	3600
tcagctgggg	gcacttcttg	aagggagggc	ctttgtgtgg	tctgtttcta	gtgacttcca	3660
gctttaaccc	agagcctcat	gattgtctgg	tgcccatagc	ctttttgtcg	aatggaggga	3720
ctcagctctc	ttgggaagag	agaatccatg	atagaccac	ttgggagctc	cccacttcag	3780
gggcctacac	actggtaatg	caacagaatg	cccaagagtg	acctcataaa	gcaaggattc	3840
ccttcgtggc	cccttctctg	ctgcctctca	gaatccagac	gctaaggaaa	atccctaagc	3900
agagattttc	tggttgatgc	taaaagcaag	gaataaaagt	tgaaaatttg	gaaaatgtct	3960
caacaccgtc	accagcgcca	ctcgagagtc	atttctagt	caccagttga	cactacatcg	4020
gtgggatttt	gcccacatt	caagaaattt	aagtaaata	tatctatctc	cattgcctgt	4080
taagaaatgt	gctagtagaa	gtgtgagggc	aggggtgtcag	tggtctctca	gcctcttccc	4140
tcagatactc	gtctgcttac	caaaataagt	tgcatgtcct	tgacaatctg	gtttctatga	4200
ttgggtgaggc	tgcatgcta	ttacctttat	gtgcctgtga	gacttgaatg	accagtttga	4260
ccagtttgac	tgtagataaa	tctgaaggct	tttctctttt	tttataatag	accccatctc	4320
aaatcagata	atgaaaatta	catatcttga	tatattagaa	aagtatatac	attctggctg	4380
ggcgcggtgg	ctcacgcctg	taatcccagc	actttgggag	gctgagggcc	ttagcagcgg	4440
ctgccgcagc	tcaatcgcg	agcaactagc	cgggcgtctg	cgggagccga	gcgtgggacc	4500
tggtgggcta	ccacctggtg	ccctcatgga	gacaagaagc	cctgggttga	acaacatgaa	4560
gccccagtc	ctgcagctgg	tactggaaga	gcaggtgctg	gcactacagc	agcagatggc	4620
agagaatcag	gcagcctcct	ggcggaagct	gaagaactcc	caggaggccc	agcagagaca	4680
agcaacccct	gtgaggaagc	tgaggccaa	ggtgctgcag	taccgaagct	ggtgccaaaga	4740
gctggagaag	cggctagaag	ccactggagg	accaatcccc	cagaggtggg	aaaatgtgga	4800
ggagccaaac	ctggatgagc	tgctgggtcg	attggaggag	gagcaacaga	ggtgtgagag	4860
tctagcacag	gtgaacaccc	agcttcgact	gcacatggaa	aaagctgacg	tggtgaataa	4920
agcccttagg	gaagatgtgg	aaaaactgac	agtggactgg	agccgggccc	gggatgagct	4980
aatgaggaag	gagagccagt	ggcagatgga	gcaggagtgg	agtctgctgt	tttcttctgt	5040
tgtagctcaga	gatctgatgg	agctaaaagc	tgagcatgtg	aggctttcag	ggtctctgtt	5100
gacctgttgt	ctgcgcttga	ctgtgggagc	acagtctcgg	gaacccaacg	gatctggaag	5160
aatgaatggg	cgggagccgg	cccagctgct	gctgctacta	gccaaagacc	aggagctgga	5220

gaaggaagcc	catgaaagga	gccaggagtt	aatacagctg	aagagtcaag	gggatctgga	5280
gaaggctgaa	cttcaggacc	gggtgaccga	gctctctgct	ctgttgaccc	agtcctcagaa	5340
gcaaaatgaa	gattatgaaa	agatgataaa	ggctctgaga	gagacagtgg	agatcctgga	5400
gacaaatcac	acagaattaa	tggaaatga	agcatctctt	agtaggaatg	cgcaagagga	5460
gaagttgtct	ttacagcagg	tgatcaagga	tataaccag	gtcatggtgg	aagaagggga	5520
caatatagcc	caaggctctg	gtcatgagaa	ctctttggaa	ttggactcta	gtatcttctc	5580
ccagtttgat	taccaggatg	cagacaaggc	tcttactctg	gtgctgtcag	tgctgactcg	5640
gagacgccag	gctgtgcagg	acctaaggca	gcagcttgca	ggctgtcaag	aggctgtgaa	5700
cttgttgcaa	cagcagcatg	atcagtggga	ggaagagggc	aaagccttga	gacagcggct	5760
gcagaagctc	actggggagc	gggacactct	ggcagggcag	actgtggacc	tccagggaga	5820
gggtggactct	ctcagcaagg	agcgagagct	gctgcagaag	gccagggaag	agctgcggca	5880
gcagctggag	gtgctagagc	aggaggcatg	gcgcctgcga	agggtaaatg	tggagcttca	5940
gctgcagggg	gactctgccc	agggccagaa	ggaggaacag	caggaggagc	tgacacctggc	6000
tgtccgggag	agggagcgtc	ttcaggagat	gctgatgggc	ctggaagcca	aacagtcaga	6060
atcactcagt	gaactgatca	ctcttcggga	agccttgga	tcaagtacc	tggaaagggga	6120
gttactgagg	caagagcaaa	cggaagtga	cgagcgcgtg	gctagggcag	agcagtcaat	6180
tgacagagctg	tcgagttctg	aaaacaccct	gaagacagaa	gtagctgatc	ttcgggctgc	6240
agctgtcaag	ctcagtgcct	taaatgaggc	tttggcggtta	gataaaggttg	ggctgaacca	6300
gcagcttctc	cagtttagagg	aggagaacca	gtctgtgtgc	agcagaatgg	aggcccgaca	6360
gcagggcaga	aatgcttttg	aaagtcgacct	ggcggaggga	gagaagagga	gggaagccct	6420
tggggaaaag	aacactcacc	tggaggctca	gctgcagaaa	gctgaggagg	ctggggctga	6480
gctgcaggca	gatctcaggg	acatccaaga	agagaaggaa	gaaattcaaa	agaaactaag	6540
tgagtcacgt	caccagcagg	aggcagccac	gactcagctg	gagcagctac	atcaggaggc	6600
aaagcgacag	gaagaagtgc	ttgccagggc	agtccaggag	aaggaggccc	tagtacgaga	6660
gaaagcggct	ctagaggtgc	ggctgcaggc	cgtggagcgt	gaccggcagg	acctcgctga	6720
acaactacag	gggctcagct	cagccaagga	gctactggag	agcagtctgt	ttgaaagcca	6780
acaacaaaat	tctgtgatag	aggtcaccaa	ggggcagctg	gaggtccaga	ttcaaactgt	6840
cactcaagcc	aaggaagtaa	tccaagggga	agtgaaggtg	ctgaagctgg	aactggacac	6900
tgaacggagt	caggcagagc	aggagcggga	tgtgcagcc	agacagctgg	cccaggctga	6960
gcaagaaggg	aagactgcct	tggagcagca	gaaggcagcc	catgagaaa	aggtgaacca	7020
ggtccgggag	aaatgggaga	aggagcgtc	ctggcaccag	caggagctgg	caaaggctct	7080
ggagagctta	gaaagggaaa	aaatggagct	ggaaatgagg	ctaaaggagc	agcagacaga	7140
aatggaggcc	atccaggccc	agagggaaga	agaacggacc	caggcagaga	gtgccctatg	7200
ccagatgcag	ctggaaacag	agaaggagag	agtatccctc	ctggagacac	tgctgcagac	7260
gcagaaggag	ctagcagatg	ccagccaaca	actggaacga	ctgaggcagg	acatgaaagt	7320
ccagaaatta	aaggagcagg	agaccactgg	gatactacag	accagctcc	aggaggtca	7380
acgggagctg	aaggaggcag	cccggcagca	cagagatgac	cttgctgccc	tccaagagca	7440
gagcagctcc	ctgctgcagg	ataagatgga	cctgcagaag	cagggtggagg	acttgaagtc	7500
tcagctggtg	gcccaggatg	actcccagag	gctggtggag	caggaggttc	aggagaagct	7560
gagagagacc	caggagtata	accgaattca	gaaggagctg	gagagagaga	aagccagcct	7620
gactctctca	ctgatggaaa	aggaacagag	actccttgtt	ttacaagaag	ctgactctat	7680
tcgacaacaa	gagctgagtg	ccctgcgcca	ggacatgcag	gaggccagg	gagaacagaa	7740
agagctcagt	gctcagatgg	aattactaag	gcaagaggtg	aaggaaaagg	aggctgactt	7800
tctggcccag	gaagcacagc	tgctggagga	gctggaggcg	tctcatatca	cggagcagca	7860
gctgcgagcc	tccttgtggg	cccagggaagc	caaggcagcc	caactacagc	tgcgactgcg	7920
cagcacagag	agccagctag	aagcgctggc	cgagagcag	cagcccggga	accaggccca	7980
tgcccaggcc	cagctggcca	gcctctactc	tgccctgcag	caggccctgg	ggtctgtttg	8040
tgagagcagg	cctgagctga	gtggtggggg	agactctgct	ccttccgtct	ggggccttga	8100
gccagaccag	aatggagcta	ggagcctctt	taagagaggg	cccctgctga	ctgctctctc	8160
cgctgaggca	gtagcatctg	ccctccacaa	gcttcatcaa	gacctgtgga	agactcaaca	8220
gaaccgggat	gttctgaggg	atcaggtcca	gaaactggaa	gagcgtctaa	ctgatactga	8280
ggctgagaag	agccaggctc	acacagagtt	gcaggatctg	cagagacagc	tctcccagaa	8340
tcaggaaagag	aaatctaagt	gggaaggaaa	gcagaactcc	ctagaatctg	agctgatgga	8400
actacatgaa	actatggcat	ccttacagag	tcgcctgcgg	agagcagagc	tacagcgaat	8460
ggaagcccag	ggtgagcgag	agttacttca	ggcagccaag	gagaacctga	cagcccagggt	8520
ggaacacctg	caagcagctg	tcgtagaagc	cagggtctcag	gcaagtgtctg	ctggcatcct	8580
ggaagaagac	ctgagaacgg	ctcgctcagc	actgaagctg	aaaaatgagg	aagtagagag	8640
tgagcgtgag	agagcccagg	ctctgcaaga	gcagggcgaa	ctgaaggtgg	cccaagggaa	8700
ggctctgcaa	gagaatttgg	ccctcctgac	ccagacccta	gctgaaagag	aagagggagt	8760
ggagactctg	cggggacaaa	tccagggaact	ggagaagcaa	cgggaaatgc	agaaggctgc	8820
tttggaattg	ctgtctctgg	acctgaagaa	gaggaaccaa	gaggtagatc	tgacagcaaga	8880

```

acagattcag gagctagaga agtgtaggctc tgttttagag catctgcccc tggccgtcca 8940
ggagcgagag cagaagctga ctgtgcagag ggagcagatc agagagctcg agaaggatcg 9000
ggagactcag aggaacgtct tggagcatca gcttctagaa cttgagaaga aagacccaaat 9060
gatttagtcc cagagaggac aggttcagga cctgaaaaag cagttggtta ctctggaatg 9120
cctggccctg gaactggagg aaaaccatca caagatggag tgccagcaaa aactgatcaa 9180
ggagctggag ggccagaggg aaaccacagag agtggctttg acccacctta cgctggacct 9240
agaagaaaag agccaggagc tgcaggcaca aagcagccag atccatgacc tggagagcca 9300
cagcaccgtt ctggcaagag agctgcagga gagggaccag gaggtgaagt ctcagcgaga 9360
acagatcgag gagctgcaga ggcagaaaaga gcatctgact caggatctcg agaggagaga 9420
ccaggagctg atgtgcaga aggagaggat tcaggttctc gaggatcaga ggaccgggca 9480
gaccaagatc ctggaggagg acctggaaca gatcaagctg tccttgagag agcgaggccg 9540
ggagctgacc actcagaggg agctgatgca ggaacgggca gaggaaggga agggcccaag 9600
taaagcacag cgcgggagcc tagagcacat gaagctgatc ctgcgtgata aggagaagga 9660
ggtggaatgt cagcaggagc atatccatga actccaggag ctcaaagacc agctggagca 9720
gcagctccag ggcctgcaca ggaaggtagg tgagaccagc ctctcctgt cccagcgaga 9780
gcaggaaata gtggtcctgc agcagcaact gcaggaagcc agggaaacaag gggagctgaa 9840
ggagcagtc cttcagagtc aactggatga ggcccagaga gccctagccc agagggacca 9900
ggaaactggag gctctgcagc aagaacagca gcaggccag ggacaggagg agaggtgaa 9960
ggaaaaggca gacgccctcc agggagctct ggagcaagcc catatgacac tgaaggagcg 10020
tcatggagag cttcaggacc acaaggaaca ggcacgaagg ctggaggaag agctggcagt 10080
ggagggacgg cgggtccagg ccctggagga ggtgctggga gacctaaagg ctgagctcg 10140
ggaacaggag aaagctctgt tggccctcca gcagcagtg gctgagcagg cacaggagca 10200
tgaggtggag accagggccc tgcaggacag ctggctgcag gccaggcag tgctcaagga 10260
acgggaccag gagctggaag ctctgcgggc agaaagtca tctcccggc atcaggagga 10320
ggctgcccgg gcccgggctg aggtctctga ggaggccctt ggcaaggctc atgtgccct 10380
gcaggggaaa gagcagcatc tcctcgagca ggcagaattg agccgcagtc tggaggccag 10440
cactgcaacc ctgcaagcct ccctggatgc ctgccaggca cacagtcggc agctggagga 10500
ggctctgagg atacaagaag gtgagatcca ggaccaggat ctccgatacc aggaggatgt 10560
gcagcagctg cagcaggcac ttgcccagag ggatgaagag ctgagacatc agcaggaacg 10620
ggagcagctg ctggagaagt ctctggccca gaggttcaa gagaatatga tccaagagaa 10680
gcagaatctg gggcaagaga gagaagagga ggagataagg ggccttcac agagtgtaa 10740
ggagctacag ctgactctag cccaaaagga acaggagatt ctggagctga gggagacca 10800
gcaaaggaa acactggaag ccttaccoca cagccacaaa acctcccaa tggaggaaca 10860
atctctaaaa cttgattctt tagagcccag gctgcagcgg gagctggagc ggctacaggc 10920
agccctgaga cagacagaag ccaggagat tgagtggagg gagaaggccc aggacttggc 10980
actctcccta gcgcagacca aggccagtgt cagcagctct caggaggtag ccatgttct 11040
acaagcctct gtcttgagc gggactcaga acagcaagg ctgcaggatg aactggagct 11100
caccagacgg gctctggaga aggagcggct acacagccca ggtgcaacca gcacagcaga 11160
actggggctc agaggggagc aggggtgtgca gctgggagag gtctcaggag tggaggctga 11220
gcctagtctt gatggaatgg agaagcagtc atggagacaa aggcttgaa acctgcagca 11280
agcagtggcc cggctggaga ttgacaggag caggctgcag cgccacaatg tccagctgcg 11340
gagtaccttg gagcaggatg ggagaggaca gaagaactca gatgccagt gtgtggctga 11400
actgcagaaa gaggtggtcc tgctgcaagc tcagctgact ttggagcggg agcagaagca 11460
ggactacatc acccgctcag cacagaccag ccgtgagcta gcaggcctgc accacagcct 11520
ctcacactca cttcttgccg tggcccaggc ccctgaggcc actgtcctgg aggcagagac 11580
ccgcaggctg gatgagctcc tgactcaaa tctgacatcc ccagggccag tcctgctaca 11640
ccccagcccc agcactacc aagcgcctc caggtag 11677

```

<210> 135  
 <211> 862  
 <212> DNA  
 <213> Homo sapiens

```

<400> 135
gttcttttcc gcaatgacta gtgtcaagct catgtactct tctgattcta gactggagaa 60
gattattcaa acttgatctg tgtttcaggt ttttaaagt cctaaaaaca gaaaattaga 120
ttcagatctc aaaaaaggaa ttttggattg actttcaaag tactaatact aattatactt 180
ttcttttggg agcgtgactc ttcttatacc taagaacata ttacaaatgt caaaaccatt 240
gcattttgac attgcaaaac atgccttgaa ctcttgaact actgtgaaaa gaatcacctg 300
tgtaaagact ttttgtaagc tagctgatac tcttaagtat gtaaaaagat tgtctttcag 360

```

```

ccgacaggcc caaaggaatg tatataagga aggaatatga aaaaataaat taggttttaa 420
aatagggaatt gggcaataaa ctgtatcaaa aatatgtaga tggatttttag tagttgtaat 480
ttaaatgtgg aaggtgaaga gaatttcaaa ctccaaagag aaatgaatga tattcagatg 540
tttcattaat ttctagtctg tgaaaatatg cattttatag taatatgtat agacttattt 600
tatttagaaa taatagtgtt ttagaattta ttaaaaaactc agtgatagcc ttatataccaa 660
aatgtttaac ttaccaaca gcaagtcata aaagtattta ttttaagct ttttaatat 720
atcgtgtaac tttcatctgt ctccagatgt aaataattat ctgcctaaat gttatatattt 780
tatgtatgca ttttctgaaa atgtattgtt ttgtaaagtg ggaaagataa taaatcaagc 840
acttcttgca cttgtttctg tg                                     862

```

<210> 136  
 <211> 1026  
 <212> DNA  
 <213> Homo sapiens

```

<400> 136
gccctttttt tctctttata caaatgaga gtatctgagc caaatatta aattctagtt 60
cttttcgca atgactagtg tcaagctcat gtactcttct gattctagac tggagaagat 120
tattcaaaat tgatctgtgt ttcagggtttt taaatgtcct aaaaacagaa aattagattc 180
agatctcaaa aaaggaatgt tggattgact ttcaaagtac taataactaat tatacttttc 240
ttttggtagc gtgactcttc ttatacctaa gaacatatta caaatgtcaa aaccattgca 300
ttttgacatt gcaaaacatg ccttgaactc ttgaactact gtgaaaagaa tcaccgttgt 360
aaagactttt tgtaagctag ctgatactct taagtatgta aaaagattgt ctttcagccg 420
acaggcccaa aggaatgtat ataaggaagg aatatgaaaa aataaattag gttttaaaat 480
aggaattggg caataaactg tatcaaaaat atgtagatgg attttagtag ttgtaattta 540
aatgtggaag gtgaagagaa tttcaaaactc caaagagaaa tgaatgatat tcagatgttt 600
cattaatttc tagtctgtga aaatatgcat tttatagtaa tatgtataga cttattttat 660
ttagaaataa tagtgtttta gaatttatta aaaactcagt gatagccttt ataccaaaat 720
gtttaacttt accaacagca agtcataaaa gtattttatt taaagctttt taatattatc 780
gtgtaacttt catctgtctt cagatgtaaa taattatctg cctaaatgtt atatttttat 840
gtatgcattt tctgaaaatg tattgttttg taaagtggga aagataataa atcaagcact 900
tcttgcaact gtttctgtga agcatataga actctatttt aaataaggac gatgtgtcgt 960
acaacaacaa atctacttgc ccgtcggttt cccggctgcg atctggccct tgccgtacca 1020
catttc                                     1026

```

<210> 137  
 <211> 611  
 <212> DNA  
 <213> Homo sapiens

```

<400> 137
gaattttaat agggccacta agaatctgag tgcttttaga gattaccctt ataccactg 60
ccatcacatc cagtcaggcc tgttgtgtc tatataaatc ttcccagctg aggggcaggt 120
gcgggctaaa atccaactgg caattggctc ccagacataa ttttatattt tacagagaag 180
catcttattg gcttatatgt gtttaaagaa tggcttggtc tatacatctt cagaaaatga 240
gaattaaaaa gtcaaaaataa ttcttgacat ctacagattg aacaaagAAC ttagaagaaa 300
taatacttta tcttttcac cttggcattcc tgagagaaga gaaattgatt gtttatcatg 360
ttggtttaat ttttcaacc agacaatctg cagcaaggca catggacccc aattttgata 420
tcgtccatc agttttcatt ctatgcatgg agctaattac tgactttgcc tgtaagaga 480
ggattgtgtg cctaaatttt gtctaacaaa tgcaagcgta gaatgacatt tactaatatt 540
tctatttctt ccataggcta aataatagta actaagtatt tttaaggaca cagccctttt 600
tttctcttta t                                     611

```

<210> 138  
 <211> 787  
 <212> DNA  
 <213> Homo sapiens

&lt;400&gt; 138

```

gagaaacaaa agctaattag ccttggggtgt ctgcttagcc ccgtgggttg agcaggagct 60
cttgacttcc ggtcttgga ctctcgccaa tagtgtgtgg aattgcactc tcacacctgc 120
tgggagacct tgcttttcaa ggggggtgtac agctcctcac cccagaaaaa tttgagatct 180
actggtttgg agggaaatgg gagataagaa tcttatgact taaggaaaac ttattttttg 240
ttatcaaat atagtgttcc ccattgaaaa tttcctttaa aactgtaaag atcttttagga 300
agtagagagt ggaagccgat gctgttctgt tcccagggtt cactcgagc tgttcccag 360
ggtcacctcc agctgttccc gagagtcacc tccagctgtt cccgagggtc acctccagct 420
ggttgggtgc atgttcttac acaagattgt aagagacgat ctatgctgtt ctgcctgtat 480
tacttgggag gcctttggct caaatgatgg aaaactctgc cccaaactca ctaatgaata 540
aagaaatgga tcatctcatg gatgagggcg tccagaggac aagggtggcc ctggggcagt 600
ggttggttgc tgctgtcatc caggacttgg gttccgtcct ctgccctctg ccaccctctg 660
tgctggcgtc aaggtggcaa ggtgtgagtt tccccgagtc acatcagctg cgacaaaatc 720
cagaggcagg aaagacctag agttcttcta aacaatgagg aaattaacat gtaaccagct 780
agaagtg

```

787

&lt;210&gt; 139

&lt;211&gt; 927

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (568) .. (738)

&lt;223&gt; a, c, g or t

&lt;400&gt; 139

```

tcaaagtgt tagtaaatat ctaaacaat gaaagggaca agatatagaa ggaatcttag 60
gatcagctga gagataattg aatactttcc taaaagaaca caatactgga agggatggg 120
ctttgtggga caattgctat tttgaattct tagtgtcca actttacaac caagggttac 180
aaatatatta aatgggtgatt tagtcagcag aagggaagac tcaaatagaa cataattagc 240
ttaagcttac ctctagttgt agagtataca ggttttgacc tcaaaatttg aaaaatcgca 300
atTTTTatct aagtgaatc aagttttcct tatttgggga tggccataat tgtctctcat 360
ggcaactcaa ctgtatgcaa cattgaacca tcctttgttg caaagggaacc tgctggggcc 420
tgcttctgt tagataaatg gtcccaactg gcataactta caagtttggg acataaacta 480
ctgattaca atactaacgc tatacgtgag caggtagtaa aatatgggca tttacatcat 540
ttattcacca agaactgtaa ttaggcannnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720
nnnnnnnnnn nnnnnnnngc tcttggctag tttctgttcc cactgggtgt tagggtttta 780
acttctgttg attcacagga tgcaggtggg ggagagatgg caagggggaa aatatgaatt 840
tgattactca agtgggaaca ggtgaaacca ttagtttaaa ataattttag gctgaaaaat 900
taaaatcatt aaattatttc cttcatg

```

927

&lt;210&gt; 140

&lt;211&gt; 4651

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 140

```

gaaacttgaa aaaaaaagac agttaatcct cattatttgt ggattctgta tttgtgaatt 60
tgctgtctcg ttaaaattta gaacacccca gtggatcctc ggtgcttctg tgggtcatgg 120
cgtgggcaga gctgcctttg ttttactgc tcagggtgaa acagtgtcct tttcgtcatt 180
tatttagtgc catgtttttt accttttttt ttttttctt ttttccctga gacagggttt 240
tgctctgtca cccaggctgg agtgcaagtgg tgtgatctca tctcactgca acctccacct 300
cccaggctta agctgtcctc ctgcctcagc cttccaagta gctgggacta cagggtgtgga 360
ccatcatgct tgggtcattt ttgtatttt ttgacatgtt acccaggctg gtcttaaaact 420
cctgggctca agtgatcctt ccgcctcggc ctcccaaagt gctgggatta cagggtgtgag 480
ctaccgcgcc cggccatttt gcacattttt gtgctttatg ttggtgattt tgggttttga 540

```



```

aatggccccg tgatgtagtg cagtctagtg ttcctgggtg caagaagggt atgacgtgca 600
gagaaaaatcc ctgtgctaga gaagcacaag ctaaggcatg agttaatgtg ctggtggcca 660
cgagttcaat gagaatgaat caatagtata tgtaaataag agtcttttaa gcacaaacag 720
acataaaata ggtttatgta ttgattgggt gacaaaaatg tcatgccccaa agactggcag 780
gaacctaaacc ctgtgtttcc cctgtgagca gtggttcagc attgactaat tctactgtct 840
tcatgacttg aaagagcata actgcagtat atgatgacga gtaaccgtaa ttcaaagtag 900
ttgatatgtg ctgagagatg ctgctgagaa agtctgggaa ctactgtttc tctggaactt 960
tcctttgtgt ggtagagggc ggagggccta tgagttcctc tctgggagtc ccttgtgtgt 1020
ctgactgtgt cccttccttc catgtgggta ggttggagcc ataatcatca cccccactcg 1080
agagtgtggc attcaaatag acgaggtcct gtgcatttc acgaagcact tccccaggtt 1140
caggtgaatt ggatgcagtg tccctgttag tcatgggctg ttttgtcgaa cttaataaaa 1200
ggctgttttc ttgtttagtc cagattcttt ggatcggagg caggaatcct ggagaagatg 1260
ttgagagggt taagcaacaa ggggtagttt gctcgtgtct gcttgtttct ttgcttgctt 1320
ttggcattga acctaagaat cgatgtgtga tctcagccta actgtgggtc cccctttcag 1380
cagcctcctc cccctgcaaa aagggtgttt ctccaggcca gcttccatt gtgctttgtc 1440
ggctccttcc cataaggccc tgatgagaac ttcacaaact caggttctta cagaggtcaa 1500
gagaacgacg tgaagataaa tgagtggta ggggctgggg gatggtgagg aactggccag 1560
cagctgcctt gaatactcag ctccagctgg tagttgcctg ggaggactcg gctcagtagt 1620
ggcagggttt gaatttttta aaaagaagat ggaattccag atttgtgtgt gtaatctcct 1680
gattttttta ctgttggcgt gtaatccaca tttgaaaaac atatcttgta ggtgagccag 1740
tctgagggag ccacatctgg ccctgagctc gccagtttgt gatctttaca taggtttctc 1800
taattctgtc aacatttcca ttttacaagt gagttacctg ggacctacaa tgggtgagta 1860
acttgcgtaa gtcagagggt ggtaagcttt tttttataa aggagcagat ggtaaatgtt 1920
ttcggtttg tgaaccatat ggtctccatc gcacttactc aactgctgtt gtagtgtgaa 1980
ggcagcgata gacaacatgt gaataaatgg gcgtggttt ctgccaataa aatacttggc 2040
aaggacaggc agcagcccag gtttggctcg ggagctggag tttgccaacc ttgcctaggt 2100
catagcccag aaccagatc ccgtctcacc acagccctca gtctttctac cctgcagcac 2160
tcttctactc tcatgctctt cgttctcttc actagggtt aaatgacttt ggggggatat 2220
aaaacttggt attttggtaa tcatcaacca cattcccaag atccagagag ttttggtaaat 2280
gttaactttc tttccaaagt gggaaacatca ttgtggccac tccaggccgc ttggaggaca 2340
tgttccggag gaaggccgaa ggcttggatc tggccagctg tgtgcgatcc ctggatgtcc 2400
tgggtgtgga tgaggcagac agacttctgg acatggggtt tgaggcaagc atcaacacca 2460
ttctggagtt tttgccaaag cagaggagaa caggccttt ctctgccact cagacgagg 2520
aagtggagaa cctggtgaga gcgggcctcc ggaacctgt ccgggtctca gtgaaggaga 2580
agggcgtggc agccagcagt gccagaaga cccctcccgc cctggaaaac tactacatgg 2640
tatgcaaggc agatgagaaa ttaatacagc tgggtccatt tcttcgcaat catagcagg 2700
agaaacacct ggtcttcttc aggcacctgt gcctgtgtgg aatactatgg gaaggctctg 2760
aagtgtctgg tgaaggcgtg tgcaattcacg gaaagatgaa atataaacgc 2820
aataagatct tcatggagtt ccgcaaatg caaagggtggg attttagtgt gcactgatgt 2880
gatggcccg ggaattgata ttctgaagt caactgggtt ttgcagtatg accctcccag 2940
caatgcaagt gccttctgtc atcgctgcgg tcgcacagct cgcattggcc acgggggag 3000
cgctctggtt ttctcctgc ccattggaaga gtcatacatc aatttcttg caattaacca 3060
aaaatgccc ctgcaggaga tgaagcccca gagaaacaca gcggaccttc tgccaaact 3120
caagtccatg gccctggctg acagagctgt gtttgaaaag ggcatgaaa cttttgtgtc 3180
atatgtccaa gcttatgcaa agcatgaatg caacctgatt ttccagattaa aggatcttga 3240
ttttgccagc cttgctcgag gttttgccct gctgaggatg cccaagatgc cagaattgag 3300
aggaagcag tttccagatt ttgtccccgt ggacgttaat accgacacga ttccatttaa 3360
agataaaatc agagaaaagc agaggcagaa actcctggag caacaaagaa gagagaaaac 3420
agaaaaatgaa gggagaagaa aattcataaa aaataaagct tgggtcaaagc agaaggccaa 3480
aaaagaaaag aagaaaaaaa tgaatgagaa aaggaaaagg gaagaggggt ctgatattga 3540
agatgaggac atggaagaac ttcttaatga cacaagactc ttgaaaaaac ttaagaaagg 3600
caaaataact gaagaagaat ttgagaaggg cttgttgaca actggcaaaa gaacaatcaa 3660
gacagtggat ttagggatct cagatttgga agatgactgc tgattccagt gccacagatg 3720
aaccacaag gacatagctg ttccctaact tgggtgatgg ctccagtttg cttttaacga 3780
aaatcacaac ttcaggagac atctgaaaag aatgatgtct ctgaaagctg tcccttcaga 3840
tgaggagaa atgaaggatt tcacacttca gaatatTTta ctaaaaacat tccagtcttg 3900
gccgggtgag gtggctcctg cctataatcc cagcactttg ggaggctgag gtgggaggat 3960
cacttgagcc caggagtcca aggtgcaggt gaataatggg tgcaccattg cactctagct 4020
tgggggacag cgtgagaccc tgtctctata ttaaatttta aaatataata aaagagaaag 4080
aaaatgccta attacagttc ttggtgaata aatgatgtaa atgccatat tttactacct 4140
gtcacgtat agcgttagta ttgtaatcag atagtttatg taccaaactt gtaagttatg 4200

```

```

ccagttggga ccatttatct aacaggaagc aggccaggc aggttccttt gcaacaaagg 4260
atggttcaat gttgcataca gttgagttgc catgagagac aattatggcc atcccaaat 4320
aaggaaaact tgattgact tagataaaaa ttgcgatttt tcaaattttg aggtcaaac 4380
ctgtatactc tacaactaga ggtaagctta agctaattat gttctatttg agtcttcct 4440
tctgctgact aaatcaccat ttaaaatatt tgtaaacctt gggtgtaaag ttggacacct 4500
aagaattcaa aatagcaatt gtccacaaa gcccatccc ttccagtatt gtgttctttt 4560
aggaaagtat tcaattatct ctgagctgat cctaagattc cttctatata ttgtcccttt 4620
catttgttta gatatttact atacactttg a                                     4651

```

&lt;210&gt; 141

&lt;211&gt; 147

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 141

```

actggcagtg cctggctcaa gaggtactca aatgtatttt aatcatttaa atgattgact 60
gcaaggtttc cagctttgag atgaatagga taatgatgtc atcatgaaca cagaagtaaa 120
tgcgtgtttg gcaataaaag agatggg                                     147

```

&lt;210&gt; 142

&lt;211&gt; 417

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (4)

&lt;223&gt; a, c, g or t

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (56)

&lt;223&gt; a, c, g or t

&lt;400&gt; 142

```

gaantaaat caattagcca gatttgtaac tttaaagcta ccacagtgag attttntctt 60
taaaaacatg catgttttta tattgtattc tttaaatatt atattgtaca gttacatgtg 120
acacatatta catatgacta atttatatgt agcaacaggg attaaagaag acctctttta 180
attgtaccac ttgttataat gctattattt acccagattc aaatgtaact acttacaatt 240
tttaaacatt tttatttctg tattctttaa aatattaagg caatggtttt tcttgaaatt 300
atattttgcc ctatgtattc tatttttatt cacactggat ttattatgat aattatatca 360
aagtgattat ttttaccctg acttttataaa gtcaactggt tcagaagtca agttgat 417

```

&lt;210&gt; 143

&lt;211&gt; 834

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 143

```

ggccctgca gggatcggcc tccagggcag gcaacagggc ctctcactga cacaggctcg 60
ggccctgcct ggccagaatg cttcttccct gggagcagcc tcggtggggc ccattgagc 120
cgcagactca agggcaggac ggagcccggt aggtcgcccc gaagcatggc gggtctcac 180
ctttgtcctg ccatggagtc agagacacga gacctggag tggggagcag gtgagtgtcc 240
tttgcctcct cctagcatcc agtaggtgcc aataaatacc tgctgactgg gtggctgaag 300
ggactgcgct ggcaccccaa ggaatttttg gatccaccac cctcgagacc gtgctgggcg 360
ggaagcatga cccacactg tgggatgggg agcctggtgg ctgcccggctg gcgcgcacc 420
caagcccttt cctgggctca gctgtgagct caggacgggt tagacgcagc agcaatgatt 480
ttaagagttg ccgtgttaaa cgggtgatagc ctgtgacggt gctgtgtgct gaaacatccg 540

```

```

atatttattgc cgctgggtgg acagcgccgt ctagacgccg ctacaccttc tcaccttcgc 600
agttgtctgt gtggaagacg ggagaggcag gggcaccacg ccaactgtccc caagaccggc 660
ccccaggcac tgcgcggaag aggcctggc ctgtgaggtg gcccttccct gccctggaca 720
ggctgtggag gggctgcagg cggaggctgg gcttgggagt gcttgaagga tcaactgggt 780
gtggaggggc ccgcaggccg cctggggatt tggggaggga agctccgggc cttt 834

```

```

<210> 144
<211> 982
<212> DNA
<213> Homo sapiens

```

```

<400> 144
ggccccctgca gggatcggcc tccagggcag gcaacagggc ctctcactga cacaggctcg 60
ggccctgacct ggccagaatg cttcttccct gggagcagcc tgggtggggc cccattgagc 120
ggcagactca agggcaggac ggagcccgtt aggtcgcccc gaagcatggc gggctctcac 180
ctttgtcctg ccatggagtc agagacacga gacctggag tggggagcag gtgagtgtcc 240
tttgctcctg cctagcatcc agtaggtgcc aataaatacc tgctgactgg gtggctgaag 300
ggactgcgct ggcatcccaa ggaattttgg gatccaccac cctcgagacc gtgctgggag 360
ggaagcatga cccacactg tgggatgggg agcctgggtg ctgcccggctg gcgcgcaccc 420
caagcccttt cctgggctca gctgtgagct caggacgggt tagacgcagc agcaatgatt 480
ttaagagttg ccgtgttaaa cgggtgtagc ctgtgacggt gctgtgtgag gaaacatccg 540
atatttattgc cgctgggtgg acagcgccgt ctagacgccg ctacaccttc tcaccttcgc 600
agttgtctgt gtggaagacg ggagaggcag gggcaccacg ccaactgtccc caagaccggc 660
ccccaggcac tgcgcggaag aggcctggc ctgtgaggtg gcccttccct gccctggaca 720
ggctgtggag gggctgcagg cggaggctgg gcttgggagt gcttgaagga tcaactgggt 780
gtggaggggc ccgcaggccg cctggggacc tggggaggga agctccgggc cttctctctg 840
ctctgcagac gcacggcctg ggttcaaatg ccagctccac cacttctctg ctgtgtggcc 900
ttggacatgt gtcttaccct ctctgagcct taaatttctt tctctggaaa atcaatgctt 960
catcaataaa aaagtgatca tg 982

```

```

<210> 145
<211> 601
<212> DNA
<213> Homo sapiens

```

```

<400> 145
aacagccttt ctgacccac agtgaggaga aactgaggca accgaagcct tgaggagggt 60
tctttcacca gtgtgagcat agctgacttt gggattcttt tgttgtgac tgtggatcaa 120
tctgtcctt gagctctcct gattttgtta taacagcaga gtccacttag attttaacca 180
tttgatcagc agagccttgt ttttggcctg ttgtgctgca tgatttacag tgtgcatctg 240
gattttgtaat tatgcctcac atagtttata ttttgattca cttcaggctt ttctgcttcc 300
tctcttttac ctgcctctag ggccatatgg agtggaggtt tgcattgtgt tgtaaaagca 360
ttttgaaatg ttcatcagac ttcatgggtt tccttacttc ctttttaggt ttgcatattt 420
aaattagggt atagaattgt gaatttacca aactatattg tggattttat tcaaacagtt 480
tctgtgggta tgaataactg gtcattgtga tgtatcaata ttttatatgg aatcattata 540
tttcttgggg ttaggctgac aaaaaacaat ctgttagctt cattcatgag aggtattcag 600
a 601

```

```

<210> 146
<211> 247
<212> DNA
<213> Homo sapiens

```

```

<400> 146
gttgatagct taatgatttc aaagaattat tatacattga aacacaaggt cagtcagtgg 60
ctggaaatga atgcacacat ttgaagttag aggactgctg agagcttgtt ggggtgggtg 120
gacctcagtc ctaggttctc tctgccccga cttctggca agggggaata gagttagctc 180
aggaaacaaa ggtaaaagcc cttttccac aactaggtaa accaatgtaa gcattaagga 240

```

tattaaa

247

&lt;210&gt; 147

&lt;211&gt; 424

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 147

```

gggggaattc agctaaaata ttaaagttgc taattctgag tcttggcaga tggctggaag 60
tccaagtcac cttttttctg agactcttgc gcatatggat gggttttggg tgggtctgct 120
atctgaagta tctgacgtgg taggtattat gggcaaatga taatgttcgt atttgaatag 180
ttgatagctt aatgatttca aagaattatt atacattgaa acacaaggtc agtcagtggc 240
tggaatatga tgcacacatt tgaagtgaga ggactgctga gagcttggtg ggtgggtggg 300
acctcagtcg taggttctct ctgccccgac cttctggcaa ggggaataga gttagctcag 360
gaaacaaagg taaaagccca ttttccacaa ctaggtaaac caagtaagca ttaaggatat 420
taaa
424

```

&lt;210&gt; 148

&lt;211&gt; 574

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 148

```

ctcgtcttgg ccctccaaat tgttgggatt acaaagtgc acatgccag ccaagagttt 60
taaaagggaa ggacatgacc agatttatgt taagagttta ccagacttc tgtgcataaa 120
atcaattagt ggaagaaaat aggtgtgat gtgacacata aggaagccag tatagaattg 180
aataagaagt aatgaagggg cttagatgag gtaatatat ttgctaccat cagagagaat 240
ggaacacaca tgagaaatat ctacaaagtg aatttgggtg aacttgggaa cagttgcata 300
tgggtggacca gtttacaagg ctgatgctca ggctgtggct tgggtgggtt gttggatggt 360
ggattaactg agacaaggag actgaccaag tagtttccct cttctctggt gctttacttt 420
cttccatagt tccttctctc tttcatagtc tagatttcta cttctcagaa aatttctctg 480
caagacggtg gaggcaaat gtgggaaggg atgttgacat caataggacc tgttagcctc 540
tgaacatctt gccaggctgg cttcagcaga ggac
574

```

&lt;210&gt; 149

&lt;211&gt; 248

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 149

```

ccactttggc taatgtaaaa aaaaaatgtg aggtgttctt aaacaaaaac ttctcaaatc 60
ctaacttaac gtaatctacc tgaaatcaat gttatgggta ctggcattga caagactgga 120
agagaaataa cttttcgtaa cagctcatct tcacatagct aatgataggc accattgttt 180
tttgggcaaa gtgatggggg aggatattaa ctatttcaaa ggttggtcaa aatgactaaa 240
atataaca
248

```

&lt;210&gt; 150

&lt;211&gt; 109

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 150

```

tttagaaagg tattcttgag cactatatga aagtaacata ttactacca attgttcttg 60
tttgaaggaa aattcctgtc ttttcagtta aaaagataag ttttgtcta
109

```

&lt;210&gt; 151

<211> 944  
 <212> DNA  
 <213> Homo sapiens

<400> 151  
 atttgacttt attcattgct gtgcaaacct ttttttgcag catatactcg ggcactacat 60  
 tcattcatca attcctgtga tgtccctggg ggtaattcaa ctcttcgagt cgcaattcat 120  
 aattttgctt ctgcacacag gcggaactttg aaaaatctat aataagaatc tgaaattaac 180  
 tggtagtatt ttggctttta cttaaaatta tccctgagag agtatttaag aaaagctgtt 240  
 caagttataa aatatataat ctgggaagaa atactgtctc atataataat tagattgtaa 300  
 tcattgtttt aatctctgtc tgggaaccaa gattgaaagc tgacttactt ctctctctcg 360  
 tcttgtgaac catacggagc ctattatattt aaaatatgat cagacaagta aggcttctct 420  
 tactttgtct tgctctgatc agaagagctc atgtgaaagc tttgagattc tcttatttat 480  
 catctttcta aaactgtgtt tttgagcttg acagtactga aaatgtcttg atgaagcaga 540  
 aaagaaagtg atgaaatgtg tttctgagca tcagagacca tctatattgc cattaccttt 600  
 tctagtgtga tataagaatt caagattaga agaatttaga tttgttgac atttttttcc 660  
 tcagcatttt tctctctgtt tttttaaaat gtattgcctc tttcccccatt cagtgcact 720  
 ggacatagga atttttaatt gtgtaatttt ctgttgcaaa aagggtaaat aaatcctttg 780  
 tcttttgaat atcttctatg tgaataaatt gtgagtga ca tttgaaaaag tgaatctgaa 840  
 gtcaaagtga gtaaatcct tatttcctta ttttttaaaa acctggtata tgacatggct 900  
 acagggcaaa tgaaataaaa attgccatag ttggtatgaa aaaa 944

<210> 152  
 <211> 3897  
 <212> DNA  
 <213> Homo sapiens

<400> 152  
 ataaggctac ggatgggagg gacggagcag cccaccgcaa agtggcggtt tacttgaggc 60  
 ggttacctta gtactccgag tagactgagt ctgtggcgag ctgcgggccg attcctggcc 120  
 agtgccatct cagccggagc aggcctcggg gcctcagaag caggctttta tctggcccca 180  
 ggctcccagc cgttcagcgc gtcttcccat aacctatacc gattattggg actctcggtc 240  
 gcagacacag gagtcacaga tgctgggaag tatggccga aagaaacctc gaaatacttc 300  
 aagggtgccc ctggctttaa accccctgaa gagcaaggac gtgttgccag tgctggctga 360  
 gaggaacgag gctatagtac cagttggggc atgggtggaa cctgcctcac caggtagttc 420  
 ggaaatccca gcatatacat cagcatattt aaftgaagaa gaactaaagg aacagctaag 480  
 aaaaaaacia gaagctttga aacattttca gaaacaagtt aaataccgag taaatcaaca 540  
 aattaggttg agaaaaaagc aacagcttca gaagtcttat gaaagagcac aaaaagaagg 600  
 ctccatagcc atgcagtctt cagcaacaca cttaacttcc aaaaggacaa gtgtttttcc 660  
 aaacaatttg aatgttgcta ttggaagttc taggttacct ccttccctga tgcctgggga 720  
 tggaaatagag gatgaagaga atcagaacga attattccaa caacaagccc aggcctctag 780  
 tgaaactatg aaacaggcac gtccacggct agcatccttt aaaaccgtga ttaaaaaaaa 840  
 gggatcagtg tttccagatg atggaaggaa aagctttctt accagagagg aagtgtttc 900  
 caggaaacca gcatccactg ggataaatac aggaataaga ggagagttgc ccattaagg 960  
 ccatcaaggc cttttagctg ctgtacctta ccagaattat atggaaaatc aggaacttga 1020  
 ctatgaggaa cctgactatg aggaatcttc atctcttgta actgatgaga aagggaaaga 1080  
 agattttgtt gggagaggcc agcaggacca gcaggctatc cattctgaag ataagaacaa 1140  
 acctttcagc agagttcaga aagtaaaatt caaaaatcca ttatttgttc tgatggaaga 1200  
 ggaagaacia aagcagttac attttgaggg ccttcaggat attctgccag aagcccagga 1260  
 ttattttcta gaagcccaag gtgatttgct ggaaacccag ggtgatttga caggaaatcca 1320  
 gagtgttaag ccagataccc aggtgttgta aatgaagggt cagggttactg agccagaagg 1380  
 ccaggccatt gagccagaag gccagcctat taagacagaa actcagggtg ttatgtctga 1440  
 agcccagagt attgagctag aagaaggagg tattgtgttg aaaaccagg attttctacc 1500  
 cacaaatcag gctcttctaa cgaaaaacca ggatgtttta ctcaaagacc actgtgttct 1560  
 ccctaaagac cagagtattc tactcaaata tcaggaccag gacttcttac ccagagacca 1620  
 gcatgttctc cacaaagacc aagatattct gccaaaatat caggaccaga attttctacc 1680  
 taaggaccag aattttttat ctagagacca gcatgttctc cccaaagacc aagatattct 1740  
 gccaaaatat caggaccaga attttctacc taaggaccag aattttttgt ctagagacca 1800  
 gcatgttctc cccaaagacc agaattattc acctaaatat caaggccagg attttctacc 1860  
 taaagaccag gactttttat ctagagacca gcatgttctc cccaaagact ggaatattct 1920

```

acccaaatgt caggaccagg attttctacc cagagaccaa ggtgttcttc ccaaagacca 1980
aaatattcta cccatatgtc aggaccagga ttttctaccc agagaccaag gttatcttcc 2040
taaagaccaa aatattctac ccatatgtca ggaccgggat tttctacca gagacctgca 2100
tggtctctcc aacgaccaga atattctacc caaatgtcag gaccaagatt ttctaccaaa 2160
atatcagaaa gtacacttta aggagccata ctctgatatg acagatgaga aagggagaga 2220
agacttttct ctggcagact atcagtgttt gcctcccaaa tcccaggacc aggatgacat 2280
caaaaatcag caacctgcat cttttatgag agaagaaaaga gtgagagagg aattgcctct 2340
ggactatcat caatatgttg tacctaaaat ccaggaccaa gactccccta gagaacagaa 2400
caagcatatc aaactaccct catcttttga gaaatgggag attgcaagag gaaatactcc 2460
tggagtacca ttggcttatg ataggtatca atcaggattg agcactgaat tccaagctcc 2520
actggcattt cagtctgacg tggataaaga agaagataag aaagagcgtc aaaagcagta 2580
cctgagacat agacgacttt tcatggatat tgagagagaa caagttaaag aacaacaaag 2640
gcaaaaagaa caaaagaaga aaattgaaaa aattaagaaa aagagagagc aagaatgtta 2700
tgctgcagag cagaggatcc taagaatgaa ctttcatgaa gatccatatt caggagagaa 2760
gttgagttag atattagccc agttacaact tcaagaaata aaaggaacca gagaaaaaca 2820
acagagagaa aaagaatacc tgagatatgt agaagcctta cgagcccaaa tccaggagaa 2880
aatgcagctg tataatatta ctttacctcc actatgctgt tgtggtcctg atttttggga 2940
tgctcatcct gatacctgtg ccaacaactg tattttctat aaaaaccaca gagcatatac 3000
tcgggcacta cattcattca tcaattcctg tgatgtccct gggggtaatt caactcttcg 3060
agtcgaatt cataattttg cttctgcaca caggcggact ttgaaaaatc tataataaga 3120
atctgaaatt aactggtagt attttggctt ttacttaaaa ttatccctga gagagtattt 3180
aagaaaagct gttcaagtta taaaatatat aatctgggaa gaaatactgt ccatataat 3240
aattagattg taatcattgt tttaatctct gtctgggaac caagattgaa agctgactta 3300
cttctctctt ctgtcttctg aaccatacgg agcctattat tttaaaatat gatcagacaa 3360
gtaaggcttc tcttactttg ctctgctctg atcagaagag ctcatgtgaa gtctttgaga 3420
ttctcttatt tatcatcttt ctaaaactgt gtttttgagc ttgacagtac tgaaaatgtc 3480
tggatgaagc agaaaagaaa gtgatgaaat gtgtttctga gcatcagaga ccatctatat 3540
tgccattacc ttttctagtt gtatataaga attcaagatt agaagaattt agatttgttg 3600
cacatttttt tcctcagcat ttttctctct tgttttttaa aatgtattgc ctctttcccc 3660
attcagtgac actggacata ggaattttta attgtgtaat tttctgttgc aaaaagggtta 3720
aataaatcct ttgtcttttg aatatcttct atgtgaaata attgtgagtg acatttgaaa 3780
aagtgaatct gaagtcaaag tgagtaaatt ccttatttcc ctatttttta aaaacctggt 3840
atatgacatg gctacagggc aaatgaaata aaaattgcc a tagttggtat gaaaaaa 3897

```

&lt;210&gt; 153

&lt;211&gt; 542

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 153

```

tagaatagaa atattaagga atctggctat ttatcctaga ttctattcca agcgactca 60
tgacagtgat ttcaagtgc cactttaaaa gaattttcat tctgaacatt taattttttt 120
gtttattcat gtttttgacc tgaacctttt tcttcgattt attattttat gtttggttac 180
ttccttctga aacggcttcc tttcattctt ttttgatttt attcctttata ttcatgtttt 240
ttccccctct ttatccaatt gagcccaagt ttgcaaaaaga aaagaacaga taaaaatatc 300
aaaattgttc atggggtagt ttttggtaat tttgtcactt tgtgtgactg gaccttctta 360
acaggcttat gataggtatc aatcaggatt gagcactgaa ttccaagctc cactggcatt 420
tcagtctgac gtggataaag aagaagataa gaaagaggt a tgtaatgata ctgcttttgg 480
atcccaatat ttctactatg atagtattaa tatatgagaa attggaaaca atttgttgtg 540
tc 542

```

&lt;210&gt; 154

&lt;211&gt; 869

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 154

```

ggtagttttt ggtaattttg tcactttgtg tgactggacc ttctaacagg cttatgatag 60
gtatcaatca ggattgagca ctgaattcca agctccactg gcatttcagt ctgacgtgga 120

```

```

taaagaagaa gataagaaag agcgtcaaaa gcagtacctg agacatagac gacttttcat 180
ggatattgag agagaacaag ttaaagaaca acaaaggcaa aaagaacaaa agaagaaaat 240
tgaaaaaatt aagaaaaaga gagagcaaga atgttatgct gcagagcaga ggatccctaag 300
aatgaacttt catgaagatc catattcagg agagaagtgg agtgagatat tagcccagtt 360
acaacttcaa gaaataaaaag gaaccagaga aaaacaacag agagaaaaag aatacctgag 420
atatgtagaa gctttacgag cccaaatcca ggagaaaatg cagctgtata atattacttt 480
acctccacta tgctgttggt gtccctgattt ttgggatgct catcctgata cctgtgcaa 540
caactgtatt ttctataaaa accacagagc atatactcgg gcactacatt cattcatcaa 600
ttcctgtgat gtccctgggg gtaattcaac tcttcgagtc gcaattcata attttgcttc 660
tgcacacagg cggactttga aaaatctata ataagaatct gaaattaact ggtagtattt 720
gggcttttac ttgaaaatca tccctgagag agtattaaga aaagctgttc aagttataaa 780
atatataatc tggaaagaaa tactgtctca tataataatt agattgtaat cattgtttta 840
atctctgtct gggaaccaag attgaaagc                                     869

```

<210> 155  
 <211> 1373  
 <212> DNA  
 <213> Homo sapiens

```

<400> 155
ggtagttttt ggtaattttg tcactttgtg tgactggacc ttctaacagg cttatgatag 60
gtatcaatca ggattgagca ctgaattcca agctccactg gcatttcagt ctgacgtgga 120
taaagaagaa gataagaaag agcgtcaaaa gcagtacctg agacatagac gacttttcat 180
ggatattgag agagaacaag ttaaagaaca acaaaggcaa aaagaacaaa agaagaaaat 240
tgaaaaaatt aagaaaaaga gagagcaaga atgttatgct gcagagcaga ggatccctaag 300
aatgaacttt catgaagatc catattcagg agagaagtgg agtgagatat tagcccagtt 360
acaacttcaa gaaataaaaag gaaccagaga aaaacaacag agagaaaaag aatacctgag 420
atatgtagaa gctttacgag cccaaatcca ggagaaaatg cagctgtata atattacttt 480
acctccacta tgctgttggt gtccctgattt ttgggatgct catcctgata cctgtgcaa 540
caactgtatt ttctataaaa accacagagc atatactcgg gcactacatt cattcatcaa 600
ttcctgtgat gtccctgggg gtaattcaac tcttcgagtc gcaattcata attttgcttc 660
tgcacacagg cggactttga aaaatctata ataagaatct gaaattaact ggtagtattt 720
gggcttttac ttaaaatcat ccctgagaga gtattaagaa aagctgttca agttataaaa 780
tatataatct ggaaagaaat actgtctcat ataataatta gattgtaatc attgttttaa 840
tctctgtctg ggaaccaaga ctgaaagctg acttacttct ctcttcttctc ttgtgaacca 900
tacggagcct attattttaa aatatgatca gacaagtaag gcttctctta ctttgctctg 960
ctctgatcag aagagctcat gtgaagctct tgagattctc ttatttatca tctttctaaa 1020
actgtgtttt tgagcttgac agtactgaaa atgtctggat gaagcagaaa agaaagtgat 1080
gaaatgtgtt tctgagcatc agagaccatc tatattgcca ttaccttttc tagttgtata 1140
taagaattca agattagaag aatttagatt tgttgacatc ttttttctctc agcatttttt 1200
cctcttggtt tttaaaatgt attgcctctt tccccattca gtgacactgg acataggaat 1260
ttttaattgt gtaattttct gttgcaaaaaa gggtaataaa atcctttgtc ttttgaatat 1320
cttctatgtg aaataaaaaa aaaaaaaaaa aaaaaaaaaa aaaattggcg gcc 1373

```

<210> 156  
 <211> 338  
 <212> DNA  
 <213> Homo sapiens

```

<400> 156
ctcctttttc aaaccgaaat ctgcagtggt tgggtgcagtt tcaagaacaa agtagtgttc 60
aacgaatact tggtgaataa atgaagggaa gaattcttga aatcattctc ccttttgctt 120
aaagattgca gtgaacttac accaaattag gcattcattg aattagctaa tattactttg 180
taaaaatata agtaccttca gtagagtcag aaactctcag gttaatagaa atgaattagt 240
ttagatttct gtcttcttaa catctaaatg aaattagctg cattgaagaa acataagttt 300
aattgagggt gcaagggtgca ggggagagtc tgggagaa                                     338

```

<210> 157

<211> 56  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> unsure  
 <222> (45)..(46)  
 <223> a, c, g or t

<400> 157  
 gggaggggat ggagagtgtc tgatggatac aggggtattt tttgnngggg cggggg 56

<210> 158  
 <211> 613  
 <212> DNA  
 <213> Homo sapiens

<400> 158  
 ttaacttcac ttaaggttag atgaagattt agttatctta tgttcaccct ttctccatac 60  
 cctccactcc caccatcact gccacatcat ctgtcctctt cccattctc ccagtattgt 120  
 aatatagttt gatcagaact cttgggttaag tctactagtgt ttacattact atggctgtgt 180  
 cactatcttc ctctgctgac ccaagcatga cactgattag gtttcctttc tcatataact 240  
 cttgtccttg gattcagtgg ccctcatttt tttcttttgc tttgttttct gttactgttc 300  
 atcacatatt ttatacagct gtcgatgtta tttattcaaa cgatgtacca gttccatttg 360  
 tttgtttgtt tctggagact ccctctggag ctttccacct tcctggctct aatctggact 420  
 ggttgcttta ggctgtggt ctaaacagct gttatttcag gacttccttg actgttctcc 480  
 tttgtttgcc ctttttttct gatctcatgg gttttttgtg tgtacttttt ttatcttgct 540  
 tacaccctca ttttgctggt acttatcttt cagtggcttt tctaagaaag attatgtagg 600  
 aagtagattt ttt 613

<210> 159  
 <211> 6239  
 <212> DNA  
 <213> Homo sapiens

<400> 159  
 tttttttttt agctgggtgt agtgggtgct gcctgtagtc ccagctacgt gagaggccga 60  
 ggtgggatga tctcttgagc ctaggaggtt gaggtgagc tgagtgtgta ttgactact 120  
 gcactctagc ctaggctata gagttagacc ctgtctcaaa aatacaata aataaattaa 180  
 ataaaagaag agggctgtcc cagttcttca gaataaccaa aggtggaatg gtgagaggac 240  
 agtatcttca tttactttct aagcattaat acctcaataa ttgggaaaaa aattaaaggc 300  
 aagaagatga aaaataagaa aaaagaggac atttaggaga tcagtccagg agggtttaag 360  
 atccaagtgc taagcacaca aattaaacaa acaaacaaaa agtcacatag aaggaatgag 420  
 aatgacatgg aaccttttaa agaaaaaaag caatgtctat aaattatgaa gaaaaataat 480  
 tttaaactca aaattatc ctagctagggt tactaatcaa aattaggggt gaataaagat 540  
 atttttaaat ataaaagaac tatgggggat ggtctcacat ttggaaaaac aaatcaaata 600  
 tacaagcagt caaaaaaatc tacttcctac ataactcttc ttagaaaagc cactgaaaga 660  
 taagtaccag caaaatgagg gtgtaagcaa gataaaaaaa gtacacacaa aaaacccatg 720  
 agatcagaaa aaaagggcaa acaaaggaga acagtcaagg aagtcctgaa ataacagctg 780  
 tttagaccac aggcctaaag caaccagtcc agattagagc caggaaagggt gaaagctcca 840  
 gagggagggt tccagaaaca acaaaacaaa tggactggt acatcgtttg aataaataac 900  
 atcgacagct gtataaaaata tgtgatgaac agtaacagaa acaaaagcaa aagaaaaaaa 960  
 tgagggccac tgaatccaag gacaagagtt atatgagaaa ggaaacctaa tcagtgtcat 1020  
 gcttggttca gcagagagaa atagtacac agccatagta atgtaaacac tagtgactta 1080  
 accaagagtt ctgatcaaac tatattcaa tactgggaga atggggaaga ggacagatga 1140  
 tgtggcagt atgggtgggag tggaggggtat ggagaaaggg tgaacataag ataactaat 1200  
 cttcatctac cttaaaatga agttaatggg cagtgtttta aattgataag aatagccat 1260  
 aagagcatat ttgaaaatat gaaggaaaaa aatagctcaa gtggttgctt ctgggcaggg 1320  
 gattagaagg gatgggagag aaaactacag tttttgttat aagcttttcc tgccatttga 1380



tttgaaactg	gggtccctccc	tctatatgtc	aaaattacaa	tttgagatga	gatttgggtg	1440
gggacacaga	gccaaaccat	atcaaactct	gtgagaaaat	ttcaaactgc	catcttagct	1500
cccagcccac	ccacactggt	tcttctacct	ggctcctttt	gctgcaggga	gctgcccag	1560
gtgtccctga	ctataccctt	gatgatgata	gcaatgagca	gacgggggct	ctgcgggagt	1620
gggtgcagcca	tagcaggatg	ctgggcattg	gggttcaaag	agttcaaagg	cgtgaagagc	1680
tgtcccagag	aggtgaccct	gagagacatc	atttattgag	agtttactcc	aagccaagca	1740
tagggctaca	tgggcttcat	atgtgttgtt	tccgatccct	acaacaacac	atcaaacttt	1800
cacgcttcga	tgaagcaagt	agccatgttg	aggaggttca	cgtggcaaaa	aagcgagaga	1860
agcttctggc	caaagccagc	aagttgctga	atctcaccaa	caacactgtg	atcttggaa	1920
cagattcttc	cccaattgag	gatcagataa	agcccagtc	tggccaacac	cttaagtgtg	1980
gtctgaagcc	gtggactgag	ctaaaccaca	ctgcatatta	cagaaactgt	gagataataa	2040
atacatgttg	tttcaagctg	ctaagtttct	gcccgtgccc	caggccatgc	cgccccatct	2100
gcgcgcggag	ccgcggctgc	cgggcctccg	gggctgagcc	gggagcgccg	ggaggaggag	2160
gcgcgcggcg	cggagcagga	gcgggagccg	cggcgggcgg	cagcgcgagg	cccagtacta	2220
tggctgtgta	ctgctatgcg	ctcaatagcc	tggatgatcat	gaatagcgcc	aacgagatga	2280
agagcgggcg	cggcccgggg	cccagtgcca	gcgagacgcc	cccgcggcgg	aggagggcag	2340
tgtctgagccc	cggcagcggt	ttcagccccc	ggagagggcg	ctctttctct	ttccccccag	2400
ccgagtcgct	gtcccccgag	gagccccgga	gccccggggg	ctggcgaggc	ggccggcgca	2460
ggctgaatag	tagcagcggc	agtggcagcg	gcagcagcgg	cagtagcgtg	agcagcccaa	2520
gcttgggtgg	tgcctgcga	ggggaccggc	agcaggtggt	ggcagccggg	accctctccc	2580
gtccaggggc	ggaggaggcc	aagaggaaag	tgcggatctt	gcagcgcgag	ttgcagaagc	2640
tgcaggtgaa	ccagaaagtg	ggcatgtttg	aggcgacat	ccaggcacag	agctccgcca	2700
ttcaagcgcc	ccgcagcccc	cgtttgggca	gggctcgctc	gccctccccg	tgccctctcc	2760
gcagcagcag	tcagccccct	ggaaggggtc	tgggtcaggg	cgcccggagc	gaggaaagg	2820
ggacaaagt	ctgggggggag	caatgtccag	agacttcagg	aaccgactcc	gggaggaaag	2880
gagggccag	cctatgctcc	tcgcaggtga	agaaaggaat	gccacctctt	cccggccggg	2940
ctgcccctac	aggatcagag	gctcagggtc	catccgcttt	tgtaggatg	gagaagggtg	3000
tccttgccag	tccccgctgt	ggctcaccca	cagctatgga	aattgacaaa	aggggctctc	3060
ctaccccggg	aactcggagc	tgcctagctc	cctcattggg	gctgttcgga	gctagcttaa	3120
cgatggccac	ggaagtggca	gcgagagtta	catccactgg	gccacaccgt	ccacaggatc	3180
ttgccctcac	tgagccgtct	gggagagccc	gtgagcttga	ggacctgcag	ccccagagg	3240
ccctgggtga	gaggcagggg	cagtttctgg	gcagtgagac	aagcccagcc	ccagaaagg	3300
gcggggcccc	cgatggagaa	ccccctggga	agatggggaa	aggatatctg	ccctgtggca	3360
tgccgggctc	tggggagcct	gaagtgggca	aaaggccaga	ggagacgact	gtgagcgtgc	3420
aaagcgcaga	gtcctctgat	tccctgagct	ggtccaggct	gcccagggcc	ctggcctccg	3480
taggcctga	ggaggcccca	agtggggccc	ccgtggggcg	ggggcggttg	cagctctccg	3540
acagagtggg	gggagggctc	ccaacgctgg	gcttgcttgg	gggcagcccc	tcagcacagc	3600
cggggaccgg	gaatgtggag	gcgggaattc	cttctggcag	aatgctggag	cctttgccct	3660
gttgggacgc	tcgaaagat	ctgaaagaac	ctcagtgcct	tcttggggac	aggggtgggtg	3720
tgcagcctgg	gaactccagg	gtttggcagg	gcaccatgga	gaaagccggt	ttggcttggg	3780
cgcgtggcac	aggggtgcaa	tcagagggga	cttgggaaag	ccagcggcag	gacagtgatg	3840
ccctcccag	tccggagctg	ctaccccag	atccggacaa	gcctttcttg	aggaaggcct	3900
gcagcccag	caacatacct	ctgtgcatca	ttacagacat	gggcacccag	gaggatgggg	3960
ccttgaggga	gacgcaggga	agccctcggg	gcaacctgcc	cctgaggaaa	ctgtcctctt	4020
cctcggcctc	ctccacgggc	ttctcctcat	cctacgaaga	ctcagaggag	gacatctcca	4080
gtgaccctga	gcgcaccctg	gaccccact	cagccttcc	gcataccctg	gaccagcaga	4140
aacctagagt	gaaatacaga	accatctgga	aggtgaagaa	caaagaaagg	gaatcaagcc	4200
ctgggaaatgc	aagtttggtg	cttattcctg	tgactgctgc	cacagggata	agggtgctag	4260
gcttgggttt	aggagacttg	ggtgaaattc	cagtctacac	atggctagca	agctctctga	4320
aaaatgggga	gagtaagtgt	gacctcatgg	agtggfactg	ctatactgtg	aagcaccag	4380
gcagcctgga	gctgcacggc	ctgcgcatga	gtccaacagg	tacaagctgt	tgtgggctta	4440
taatgtcagc	acctaagcag	gaattgaatg	caatcgagtt	gagttacctg	cctccagctc	4500
ccatagtgtg	ggtgaggaaa	agtggcttca	gtgctcagca	gagtgttggg	gactgcatca	4560
agcccagttc	gcctatcagg	gacagggtag	ctctcctgtg	ccctatgggc	ttcaaggcaa	4620
aagggcttta	tgaatcctgc	ctctggcaca	gcccagagtc	cagtggatc	cggcagaagc	4680
agtgtgtg	agctctgagc	tgggcccctt	aggggaagag	ggagtacctc	cagcaatatt	4740
caggggtgat	gtgggttcca	gggcttctga	tcctggggct	gggcttgtct	gagattcaca	4800
ggagctcctt	gcaagttcag	cctgctggag	gtgtacacac	agaagcagca	gcccctggag	4860
ccctggaca	ccagggagcc	atgtccgtca	catatgatgc	cttgaggag	aaacagcagc	4920
tcagcaaaagt	aggtgacttg	cccgcgttga	catggccagg	tcctcttatc	agccagatgc	4980
cgggagtatt	ggattcctgc	cggctgtgca	gcctggggga	catagagaag	agcaaatcat	5040

```

ggaggaagat aaaaaacatg gtgcactggt ctcccttcgt catgtccttc aagaagaagt 5100
acccttgat ccagctggca ggacacgcag ggagtttcaa ggcagctgcc aatggcagga 5160
tcctgaagaa gcactgtgag tcagagcagc gctgcctgga ccggtgatg gtggatgtgc 5220
tgaggccctt cgtacctgcc taccatgggg atgtggtgaa ggacggggag cgctacaacc 5280
agatggacga cctgctggcc gacttcgact cgccctgtgt gatggactgc aagatgggaa 5340
tcagacagca gcaggacttc gcaggtgacc acatggagaa taatccaagt ggtgtccact 5400
cagacctggc caaaaaagca ggggagtggt gggaggggct gagcctcacc ttctgtggg 5460
catcccgccc caccatccag ctggcaccce ctgtggacat ctccccccag cctctttcct 5520
ctcctgggca gacctacctg gaggaggagc tcacgaaggc ccggaagaag cccagcctgc 5580
ggaaggacat gtaccagaag atgatcgagg tggaccccgaggccccacc gaggaggaaa 5640
aagcacagcg ggctgtgacc aagccacggt acatgcagt gcgggagacc atcagctcca 5700
cggccaccct ggggttcagg atcgaggga tcaagctaag aggcctctgcc tggggtgcac 5760
tgcccacagc ccccggtctt cgccccctcc tgcacccagg gctgctccct cagccccagg 5820
tcctgccagt cctgtcgaag gcagccacaa aagaagacgg caccgtgaac cgggacttca 5880
agaagaccaa aacgagggag caggtcaccg aggccttcag agagttcact aaaggaaacc 5940
ataacatcct gatcgctat cgggaccggc tgaaggccat tcgaaccact ctagaagttt 6000
ctcccttctt caagtgccac gaggtcattg gcagctccct cctcttcac cagacaaga 6060
aggaacagcg caaagtgtgg atgatcgact ttgggaaaac cagccccctg cctgagggcc 6120
agacctgca gcatgacgtc ccctggcagg aggggaaccg ggaggatggc tacctctcgg 6180
ggctcaataa cctcgtcgac atcctgaccg agatgtccca ggatgcccc ctcgctga 6239

```

<210> 160  
 <211> 15  
 <212> PRT  
 <213> Homo sapiens

<400> 160  
 Met Asn Asn Ser Gly Ala Asp Leu His Leu Ser Thr Gly Thr Ile  
 1 5 10 15

<210> 161  
 <211> 37  
 <212> PRT  
 <213> Homo sapiens

<400> 161  
 Met Asn Tyr Lys Leu Ser Glu Ile Ile Leu Ser Ser Lys Leu Ile Thr  
 1 5 10 15

Asp Val Ser Glu Ile Thr Gln Ile Met Phe Pro Phe Gln Phe Lys Ser  
 20 25 30

Arg Pro Phe Pro Leu  
 35

<210> 162  
 <211> 94  
 <212> PRT  
 <213> Homo sapiens

<400> 162  
 Met Gly Gln Glu Ala Gly Val Trp Gln Val Ser Phe Cys Phe Lys Lys  
 1 5 10 15

Gly Lys Gln Lys Glu Cys Gln Lys Phe Asp Phe Asn Phe Leu Ala Glu  
 20 25 30

Ala Phe Leu Pro Phe Ser Cys Pro Phe Phe Phe Pro Leu Pro Ser Phe

35

40

45

Pro Pro Ser Val Leu Ser Ser Phe Leu Phe Pro Leu Leu Ile Pro Phe  
 50 55 60

His Arg Thr Phe Cys Ala Gln Lys Met Thr Ala Ser Cys His Ala Pro  
 65 70 75 80

Leu Cys Glu Ser Ser Cys Ser Leu His Cys Gln Leu His Phe  
 85 90

&lt;210&gt; 163

&lt;211&gt; 53

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 163

Met Thr Leu Asn Glu His Ala Ala Phe Lys His Leu Phe Asn Glu Ala  
 1 5 10 15

His Leu Ala Pro Pro Leu Ile His Leu Thr Leu Ser Gly His Ser Thr  
 20 25 30

Cys Phe Arg Glu His Arg Val Gly Gly Thr Val Pro Asp Thr Gly Asp  
 35 40 45

Asn Lys Glu Lys Gln  
 50

&lt;210&gt; 164

&lt;211&gt; 31

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 164

Met Leu Ile Cys Phe Tyr Pro Asp Thr Tyr Asn Gln Val Glu Leu Gly  
 1 5 10 15

Ile Leu Phe Ser Leu Arg Val Gly Glu His Arg Ile Thr Leu Tyr  
 20 25 30

&lt;210&gt; 165

&lt;211&gt; 36

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 165

Met Ile Thr Lys Ile Ile Asn Tyr Leu Gln Ile Ile Phe Thr Gly Ile  
 1 5 10 15

Val Arg Pro Ile Arg Lys Asn Tyr Lys Thr Leu Trp Asp Gly Tyr Lys  
 20 25 30

Arg Arg Phe Glu  
 35

&lt;210&gt; 166

<211> 19  
 <212> PRT  
 <213> Homo sapiens

<400> 166  
 Met Phe Leu Asn Cys Thr Met Asn Tyr Lys Asn Leu Leu Ala Arg Ser  
           1                          5                          10                          15

Val Leu Phe

<210> 167  
 <211> 22  
 <212> PRT  
 <213> Homo sapiens

<400> 167  
 Met Lys Cys Phe Ser Phe Cys Leu Asn Thr Thr Ser Phe Thr Val Val  
           1                          5                          10                          15

Lys Val Asn Tyr Phe Pro  
                           20

<210> 168  
 <211> 68  
 <212> PRT  
 <213> Homo sapiens

<400> 168  
 Met Arg Leu Phe Ala Ile Val Gly Cys Trp Lys Phe Gly Tyr Ser Lys  
           1                          5                          10                          15

Trp Tyr Ile Arg Leu Leu Phe Ala Cys Ala Pro Glu Val Phe Val Pro  
                           20                          25                          30

Ala Ser Arg Ser Ala Val Ser Thr Pro Leu Ser Gln Pro Val Gly Ser  
                           35                          40                          45

Thr Cys Glu Lys Leu Ser Ile Pro Gly Leu Ser Gly Arg Phe Leu Thr  
           50                          55                          60

Ser Leu Met Phe  
           65

<210> 169  
 <211> 105  
 <212> PRT  
 <213> Homo sapiens

<400> 169  
 Phe Leu Leu Arg Gln Asp Leu Thr Leu Ser Pro Lys Leu Glu Cys Ser  
           1                          5                          10                          15

Gly Ala Ile Met Ala His Cys Ser Leu Gly Leu Pro Gly Ser Ser Asn  
                           20                          25                          30

Pro Ser Thr Ser Ala Ser Arg Leu Ala Gly Thr Thr Gly Ala Tyr His  
                           35                          40                          45

Gln Ala Trp Leu Ile Phe Leu Ile Lys Thr Gly Val Tyr Tyr Val Ala  
 50 55 60

Gln Ala Gly Leu Glu Leu Leu Asp Ser Ser Asn Ser Pro Thr Leu Ala  
 65 70 75 80

Ser Gln Ser Asp Arg Ile Thr Gly Met Ser His His Ala Gln Pro Gly  
 85 90 95

Ser Pro Leu Leu Thr Ile Thr Ile Pro  
 100 105

<210> 170  
 <211> 35  
 <212> PRT  
 <213> Homo sapiens

<400> 170  
 Met Leu Thr Ile Ser Glu Lys Ile Ile Ser Tyr Ile Tyr Ile Leu Val  
 1 5 10 15

Ser Lys Asp Ala Leu Lys Ala Leu Ser Ser Ile Val His Asn Ile Pro  
 20 25 30

Gly Leu Phe  
 35

<210> 171  
 <211> 78  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> UNSURE  
 <222> (74)

<400> 171  
 Met Ala Leu Gly His Ile Ser Gln Trp Ser Asp Pro Gly Ser Gln Gln  
 1 5 10 15

Ser Leu Leu Ser Ile Arg Asp Arg Thr Met Ala Gly Thr Leu Ser Lys  
 20 25 30

Val Pro His Asp Pro Glu Asp Met Cys Glu Phe Cys Ile Ile Phe Pro  
 35 40 45

Ser Ile Ile Leu Arg Thr Val Arg Ala Lys Val Arg Thr Leu Thr His  
 50 55 60

Arg Phe Val Thr Arg Arg Asn Ser Leu Xaa Thr Glu Ser Phe  
 65 70 75

<210> 172  
 <211> 32  
 <212> PRT  
 <213> Homo sapiens

&lt;400&gt; 172

Met Arg Pro Gly Trp Pro Leu His Phe Leu Arg Asp Val Met Asn Ser  
 1 5 10 15  
 Arg Val Thr Lys Met Gln Thr Ala Ser Ser Arg His Arg Gly Met Val  
 20 25 30

&lt;210&gt; 173

&lt;211&gt; 46

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 173

Met Glu Lys Asp Leu Arg Val Gln Ser Ser Gly Pro Ile Leu Pro Arg  
 1 5 10 15  
 Arg Leu Gly Lys Phe Met Arg Val Ser Gly Arg Gly His Gly Val Leu  
 20 25 30  
 Ile Asp Leu Phe Ser Gln Leu Lys Ser Ser Phe Arg Leu Ser  
 35 40 45

&lt;210&gt; 174

&lt;211&gt; 39

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 174

Met Val Cys Arg Cys Ser Arg Lys Leu Cys Arg Trp Tyr Val Gly Asn  
 1 5 10 15  
 Trp Ile Trp Gly Asn Ala Ala Ala Cys His Ala Leu Ser Ile Gly Arg  
 20 25 30  
 Phe Ser Pro Leu Phe Pro Pro  
 35

&lt;210&gt; 175

&lt;211&gt; 38

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 175

Met Asn Thr Thr Leu Leu Cys Leu Cys Arg Ile Leu Pro Glu His Gly  
 1 5 10 15  
 Gly Lys Ser Thr Gly Ile Val Val Arg Lys Leu Gly Phe Trp Pro Glu  
 20 25 30  
 Phe Ala Pro Asp Tyr Gln  
 35

&lt;210&gt; 176

&lt;211&gt; 36

<212> PRT  
<213> Homo sapiens

<400> 176  
Met Leu Ala Lys Ile Ser Lys Thr Ile Lys Pro Gly Ser Ile Glu Leu  
1 5 10 15  
Pro Ser Ser Tyr His Lys Val Phe Pro His Phe Leu Leu Ile Val Asn  
20 25 30  
Phe Leu Lys Lys  
35

<210> 177  
<211> 51  
<212> PRT  
<213> Homo sapiens

<220>  
<221> UNSURE  
<222> (26)..(32)

<400> 177  
Met Phe Ser Ser Pro Ser Asp Cys Leu Leu Ile Pro His Leu Phe Phe  
1 5 10 15  
Arg Ser Leu Phe Phe Ile His Trp Leu Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
20 25 30  
Ala Phe Lys Phe Leu Leu Phe Met Arg Gln Met Tyr Leu Arg Ser Ile  
35 40 45  
Asp Val Ser  
50

<210> 178  
<211> 15  
<212> PRT  
<213> Homo sapiens

<400> 178  
Met Leu Ala Asn Thr Ile Val Ser Val Arg Lys Cys Arg Val Trp  
1 5 10 15

<210> 179  
<211> 57  
<212> PRT  
<213> Homo sapiens

<400> 179  
Met Ser Ser Leu Leu Lys Ala Leu Thr Phe Trp Pro Gln Arg Met Ala  
1 5 10 15  
Leu Phe Val Pro Ile Arg Thr Arg Ile Leu Ile Phe Leu Leu Gly  
20 25 30  
Pro Gly Asn Gln Arg Thr Thr Asn Thr Phe Ala Arg His Leu Gln Pro  
35 40 45

Ser Arg Ser Gly Arg Pro Ser Leu Ser  
50 55

<210> 180  
<211> 46  
<212> PRT  
<213> Homo sapiens

<400> 180  
Met Arg Asn Ile Asn Ile Val Asp Tyr Ile Lys Ile Gly Ser Phe Cys  
1 5 10 15

Ser Ser Thr Met Ser Glu Gly Glu Lys Ala Ser His Ile His His Pro  
20 25 30

Tyr Ala Pro Lys Thr Gly Met Pro Arg Ala Glu Phe Arg Ala  
35 40 45

<210> 181  
<211> 47  
<212> PRT  
<213> Homo sapiens

<220>  
<221> UNSURE  
<222> (26)..(47)

<400> 181  
Met Leu Asn Met Pro Leu Thr Ile Gln Ile Met Tyr Tyr Leu Met Leu  
1 5 10 15

Leu Ile Ile Val Leu Phe Asn Leu Arg Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
20 25 30

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
35 40 45

<210> 182  
<211> 45  
<212> PRT  
<213> Homo sapiens

<400> 182  
Met Ser Thr Ile Arg Glu His Ile Ser Leu Tyr Ile Ile Val Thr Asn  
1 5 10 15

Ile Leu Asn Tyr Lys Glu Lys Lys Lys Lys Asp Ala Lys Val Gln Arg  
20 25 30

Leu Asn Ser Gln His Pro Thr Asp Arg Glu Tyr Leu Gly  
35 40 45

<210> 183  
<211> 57  
<212> PRT  
<213> Homo sapiens



&lt;400&gt; 183

Met Phe Cys Val Tyr Val Lys Pro Ser Pro Pro Val Leu Phe Ile Gly  
 1 5 10 15

Gly Gly Leu Ile Ala Val Met Ala Ser Ile Asn Gly Phe Leu Val Pro  
 20 25 30

Arg Pro Ser Val Val Leu Ser His Ser Asp Ser Arg Leu Asn Asn Met  
 35 40 45

Ala Lys Glu Glu Ser Arg Lys Leu Glu  
 50 55

&lt;210&gt; 184

&lt;211&gt; 28

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 184

Met Leu Ile Phe Leu Phe Tyr Ser Ile Pro Ile Ser Arg Ala Gln Leu  
 1 5 10 15

Ile Gly Gln Pro Thr Thr Gly Ser Pro Cys Trp Val  
 20 25

&lt;210&gt; 185

&lt;211&gt; 27

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 185

Met Pro Thr Arg Val Phe Ile Thr His Tyr Tyr Ser Ile Phe Gly Val  
 1 5 10 15

Pro Val Pro Cys Ser Leu Asn Asn Pro Gln Leu  
 20 25

&lt;210&gt; 186

&lt;211&gt; 25

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 186

Met Gln Arg Gly Lys Glu Leu Ile Val Ala Leu Phe Glu Asn Tyr Leu  
 1 5 10 15

Arg Pro Ser Leu Gly His Phe Asn Ser  
 20 25

&lt;210&gt; 187

&lt;211&gt; 49

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 187

Met Leu Ser Gln Phe Leu Lys Met Glu Trp Glu Val Glu Ile Ser Gln

1                      5                      10                      15  
 Val Val Ala Gly Leu Gln His Phe His Ile Leu Gly Tyr Ile Ile Thr  
                             20                      25                      30  
 Arg Cys Cys Leu Pro Ala Gly Ala Ile Thr Ala Ser Lys Ala Thr Cys  
                             35                      40                      45  
 Phe

<210> 188  
 <211> 113  
 <212> PRT  
 <213> Homo sapiens

<400> 188  
 Met Ala Thr Lys Gln Ser Pro Leu Phe Tyr Leu Thr Gly Ser Ala Gly  
   1                      5                      10                      15  
 Gly Ser Leu Val Leu Lys Pro Pro Pro Asn His Pro Tyr Arg Val Ser  
                             20                      25                      30  
 Leu Arg Ala Lys Met Met Pro Gln His Pro Arg Arg Pro Leu Leu Pro  
                             35                      40                      45  
 His Gln Leu Gly Thr Lys Tyr Ser Leu Lys Cys Phe Ala Cys Gln Thr  
                             50                      55                      60  
 Thr Arg Lys Gly Asn Ala Val Ser Thr Ser Ser Ile Cys Leu Cys Leu  
                             65                      70                      75                      80  
 Val Arg Arg Ala Leu Glu Glu Phe Arg Met Gln Val Lys Ser Met Glu  
                             85                      90                      95  
 Gly Gly Ile Ser Phe Leu Ile Cys Lys Met Ser Leu Ile Lys Leu Ile  
                             100                      105                      110  
 Thr

<210> 189  
 <211> 31  
 <212> PRT  
 <213> Homo sapiens

<400> 189  
 Met Pro Gln Thr Cys Thr Tyr Ser Lys Ser Asn Ile Leu Lys Ile Tyr  
   1                      5                      10                      15  
 Gly Ile Asp Arg Asn Thr Phe Lys Ala Thr Ile His Thr Ala Arg  
                             20                      25                      30

<210> 190  
 <211> 38  
 <212> PRT  
 <213> Homo sapiens

&lt;400&gt; 190

Met Gln Phe Gln Ala Leu Gly Arg Arg Val Pro Asp Cys Phe Leu Tyr  
 1 5 10 15

Thr Ala Ile Ile Pro Tyr Thr Ala Gly Ser Ser Phe Phe Asp Ile Leu  
 20 25 30

Cys Asn Cys Arg Gly Leu  
 35

&lt;210&gt; 191

&lt;211&gt; 78

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 191

Met Lys Ile Pro Ala Leu Ser Trp Val Trp Pro Ser Arg Asn Leu Leu  
 1 5 10 15

Ser Tyr Ile His Gly Val Leu Pro Phe Tyr Lys Leu Met Phe Cys Asn  
 20 25 30

His Pro Gly Tyr Phe Pro Arg Arg Lys Lys Lys Leu Val Glu Gln Gly  
 35 40 45

Glu Gly Cys Leu Lys Phe Gly Asn His Pro Trp Tyr Leu Asn Gln Gly  
 50 55 60

Lys Ala Leu Arg Ser Leu Val Leu Gly Asn Ile Leu Leu Tyr  
 65 70 75

&lt;210&gt; 192

&lt;211&gt; 34

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 192

Met Leu His Val Cys Ser Val Leu Ser Arg Gln Arg Leu Ala Pro Met  
 1 5 10 15

Lys Glu Ala Ser Glu Pro Ser Arg Arg Glu Val Phe Ser Leu Ser Asn  
 20 25 30

Ser Gln

&lt;210&gt; 193

&lt;211&gt; 325

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 193

Lys Val Ser Ile Leu Ser Thr Phe Leu Ala Pro Phe Lys His Leu Ser  
 1 5 10 15

Pro Gly Ile Thr Asn Thr Glu Asp Asp Asp Thr Leu Ser Thr Ser Ser  
 20 25 30

Ala Glu Val Lys Glu Asn Arg Asn Val Gly Asn Leu Ala Ala Arg Pro  
                   35                                  40                                  45  
 Pro Pro Ser Gly Asp Arg Ala Arg Gly Gly Ala Pro Gly Ala Lys Arg  
                   50                                  55                                  60  
 Lys Arg Pro Leu Glu Glu Gly Asn Gly Gly His Leu Cys Lys Leu Gln  
                   65                                  70                                  75                                  80  
 Leu Val Trp Lys Lys Leu Ser Trp Ser Val Ala Pro Lys Asn Ala Leu  
                                   85                                  90                                  95  
 Val Gln Leu His Glu Leu Arg Pro Gly Leu Gln Tyr Arg Thr Val Ser  
                                   100                                  105                                  110  
 Gln Thr Gly Pro Val His Ala Pro Val Phe Ala Val Ala Val Glu Val  
                                   115                                  120                                  125  
 Asn Gly Leu Thr Phe Glu Gly Thr Gly Pro Thr Lys Lys Lys Ala Lys  
                   130                                  135                                  140  
 Met Arg Ala Ala Glu Leu Ala Leu Arg Ser Phe Val Gln Phe Pro Asn  
                   145                                  150                                  155                                  160  
 Ala Cys Gln Ala His Leu Ala Met Gly Gly Gly Pro Gly Pro Gly Thr  
                                   165                                  170                                  175  
 Asp Phe Thr Ser Asp Gln Ala Asp Phe Pro Asp Thr Leu Phe Gln Glu  
                                   180                                  185                                  190  
 Phe Glu Pro Pro Ala Pro Arg Pro Gly Leu Ala Gly Gly Arg Pro Gly  
                                   195                                  200                                  205  
 Asp Ala Ala Leu Leu Ser Ala Ala Tyr Gly Arg Arg Arg Leu Leu Cys  
                   210                                  215                                  220  
 Arg Ala Leu Asp Leu Val Gly Pro Thr Pro Ala Thr Pro Ala Ala Pro  
                   225                                  230                                  235                                  240  
 Gly Glu Arg Asn Pro Val Val Leu Leu Asn Arg Leu Arg Ala Gly Leu  
                                   245                                  250                                  255  
 Arg Tyr Val Cys Leu Ala Glu Pro Ala Glu Arg Arg Ala Arg Ser Phe  
                                   260                                  265                                  270  
 Val Met Ala Val Ser Val Asp Gly Arg Thr Phe Glu Gly Ser Gly Arg  
                                   275                                  280                                  285  
 Ser Lys Lys Leu Ala Arg Gly Gln Ala Ala Gln Ala Ala Leu Gln Glu  
                   290                                  295                                  300  
 Leu Phe Asp Ile Gln Met Pro Gly His Ala Pro Gly Arg Ala Arg Arg  
                   305                                  310                                  315                                  320  
 Thr Pro Met Pro Gln  
                                   325

&lt;210&gt; 194

&lt;211&gt; 33

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 194

Met Ala Ser Phe Leu Leu Ser Thr Pro Ala Lys Arg Lys Pro His Pro  
 1 5 10 15

Leu Pro Pro Ala His Pro Arg Ile His Thr Phe Arg Gln Pro Ser Gly  
 20 25 30

Asn

&lt;210&gt; 195

&lt;211&gt; 74

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 195

Met Ile Pro Thr Phe Val Leu Asp Ala Lys Tyr Ala Ala Leu Met Gly  
 1 5 10 15

Gln Pro Trp Gly Leu Cys Ala Ile Cys Val His Ile Cys Leu Leu Leu  
 20 25 30

Asp Ser Val Ser Leu Arg Ser Phe Ser Thr Ala Gln His Leu Glu Arg  
 35 40 45

Ala Ser Lys Ser Thr Ser Ser Leu His His Leu Ile Leu Ile Asn Pro  
 50 55 60

Ala Arg Glu Gly Cys Thr Gly Arg Thr Ala  
 65 70

&lt;210&gt; 196

&lt;211&gt; 97

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 196

Met Ala Asn Phe Cys Val Phe Ile Glu Thr Glu Gly Asn Ala Val Thr  
 1 5 10 15

Arg Arg Ala Leu Arg Lys Gln Ala Thr Ala Gly His Cys Ser Gly Lys  
 20 25 30

Pro Ala Phe Gln Pro Ala Pro Pro Gln Tyr Pro Arg Val His Ser Glu  
 35 40 45

Asp Arg Arg Leu Gln Gln Pro Gln Ala Ala Gly Arg Trp Gly Ala Pro  
 50 55 60

Asp Trp Ile Pro Pro Leu Gln Asp Thr Arg Lys Pro Ser Val Ser Ser  
 65 70 75 80

Arg Asp Ser Arg Ile His Glu Lys Glu Val Ile Leu Asp Ser Leu Cys  
 85 90 95

Ile

<210> 197  
 <211> 645  
 <212> PRT  
 <213> Homo sapiens

<400> 197  
 Ala Leu Arg Pro Pro Ser Gly Phe His Ile Arg Cys Leu Gly Asp Val  
           1                  5                  10                  15  
 Ser Pro Ile Ser Met Ser Pro Ile Ser Gln Ser Gln Phe Ile Pro Leu  
                   20                  25                  30  
 Gly Glu Ile Leu Cys Leu Ala Ile Ser Ala Met Asn Ser Ala Arg Lys  
                   35                  40                  45  
 Pro Val Thr Gln Glu Ala Leu Met Glu His Leu Thr Thr Cys Phe Pro  
                   50                  55                  60  
 Gly Val Pro Thr Pro Ser Gln Glu Ile Leu Arg His Thr Leu Asn Thr  
                   65                  70                  75                  80  
 Leu Val Arg Glu Arg Lys Ile Tyr Pro Thr Pro Asp Gly Tyr Phe Ile  
                           85                  90                  95  
 Val Thr Pro Gln Thr Tyr Phe Ile Thr Pro Ser Leu Ile Arg Thr Asn  
                   100                  105                  110  
 Ser Lys Trp Tyr His Leu Asp Glu Arg Ile Pro Asp Arg Ser Gln Cys  
                   115                  120                  125  
 Thr Ser Pro Gln Pro Gly Thr Ile Thr Pro Ser Ala Ser Gly Cys Val  
                   130                  135                  140  
 Arg Glu Arg Thr Leu Pro Arg Asn His Cys Asp Ser Cys His Cys Cys  
                   145                  150                  155                  160  
 Arg Glu Asp Val His Ser Thr His Ala Pro Thr Leu Gln Arg Lys Ser  
                           165                  170                  175  
 Ala Lys Asp Cys Lys Asp Pro Tyr Cys Pro Pro Ser Leu Cys Gln Val  
                   180                  185                  190  
 Pro Pro Thr Glu Lys Ser Lys Ser Thr Val Asn Phe Ser Tyr Lys Thr  
                   195                  200                  205  
 Glu Thr Leu Ser Lys Pro Lys Asp Ser Glu Lys Gln Ser Lys Lys Phe  
                   210                  215                  220  
 Gly Leu Lys Leu Phe Arg Leu Ser Phe Lys Lys Asp Lys Thr Lys Gln  
                   225                  230                  235                  240  
 Leu Ala Asn Phe Ser Ala Gln Phe Pro Pro Glu Glu Trp Pro Leu Arg  
                           245                  250                  255  
 Asp Glu Asp Thr Pro Ala Thr Ile Pro Arg Glu Val Glu Met Glu Ile  
                   260                  265                  270  
 Ile Arg Arg Ile Asn Pro Asp Leu Thr Val Glu Asn Val Met Arg His  
                   275                  280                  285

Thr Ala Leu Met Lys Lys Leu Glu Glu Glu Lys Ala Gln Arg Ser Lys  
 290 295 300  
 Ala Gly Ser Ser Ala His His Ser Gly Arg Ser Lys Lys Ser Arg Thr  
 305 310 315 320  
 His Arg Lys Ser His Gly Lys Ser Arg Ser His Ser Lys Thr Arg Val  
 325 330 335  
 Ser Lys Gly Asp Pro Ser Asp Gly Ser His Leu Asp Ile Pro Ala Glu  
 340 345 350  
 Arg Glu Tyr Asp Phe Cys Asp Pro Leu Thr Arg Arg Ser Asn Lys Ala  
 355 360 365  
 Lys Glu Arg Ser Arg Ser Met Asp Asn Ser Lys Gly Pro Leu Gly Ala  
 370 375 380  
 Ser Ser Leu Gly Thr Pro Glu Asp Leu Ala Glu Gly Cys Ser Gln Asp  
 385 390 395 400  
 Asp Gln Thr Pro Ser Gln Ser Tyr Ile Asp Asp Ser Thr Leu Arg Pro  
 405 410 415  
 Ala Gln Thr Val Ser Leu Gln Arg Ala His Ile Ser Ser Thr Ser Tyr  
 420 425 430  
 Lys Glu Val Cys Ile Pro Glu Ile Val Ser Gly Ser Lys Glu Pro Ser  
 435 440 445  
 Ser Ala Cys Ser Leu Leu Glu Pro Gly Lys Pro Pro Glu Ser Leu Pro  
 450 455 460  
 Ser Tyr Gly Glu Leu Asn Ser Cys Pro Thr Lys Thr Ala Thr Asp Asp  
 465 470 475 480  
 Tyr Phe Gln Cys Asn Thr Ser Thr Tyr His Lys Ser Ser Leu Ser Leu  
 485 490 495  
 Leu Lys Ser His Pro Lys Thr Pro Ala Asp Thr Leu Pro Gly Arg Cys  
 500 505 510  
 Glu Lys Leu Glu Pro Ser Leu Gly Thr Ser Ala Ala Gln Ala Met Pro  
 515 520 525  
 Ala Ser Gln Arg Gln Gln Glu Ser Gly Gly Asn Gln Glu Ala Ser Phe  
 530 535 540  
 Asp Tyr Tyr Asn Val Ser Asp Asp Asp Asp Ser Glu Glu Gly Ala Asn  
 545 550 555 560  
 Lys Asn Thr Glu Glu Glu Lys Asn Arg Glu Asp Val Gly Thr Met Gln  
 565 570 575  
 Trp Leu Leu Glu Arg Glu Lys Glu Arg Asp Leu Gln Arg Lys Phe Glu  
 580 585 590  
 Lys Asn Leu Thr Leu Leu Ala Pro Lys Glu Thr Asp Ser Ser Ser Asn  
 595 600 605

Gln Arg Ala Thr His Ser Ala Arg Leu Asp Ser Met Asp Ser Ser Ser  
 610 615 620

Ile Thr Val Asp Ser Gly Phe Asn Ser Pro Arg Cys Pro Ala Ala Leu  
 625 630 635 640

Lys Ala Glu Ala Ser  
 645

<210> 198  
 <211> 29  
 <212> PRT  
 <213> Homo sapiens

<400> 198  
 Met Leu Leu Tyr Ser Thr Arg Gly Lys Lys His Gly Leu Tyr Pro Gln  
 1 5 10 15

Gln Ser Leu Gly Asn Arg Gly Ile Tyr Leu Gln Asn Gly  
 20 25

<210> 199  
 <211> 32  
 <212> PRT  
 <213> Homo sapiens

<400> 199  
 Met Val Thr Lys Lys Asn Leu Lys Ser Asn Asn Leu Val Gly Ala His  
 1 5 10 15

Leu Glu Tyr Asn Ser Met Ser Ser Cys Ile Tyr Leu Ser His Ile Leu  
 20 25 30

<210> 200  
 <211> 38  
 <212> PRT  
 <213> Homo sapiens

<400> 200  
 Leu Ala Asn Phe Arg Ile Phe Ser Arg Asp Arg Val Ser Pro Cys Trp  
 1 5 10 15

Pro Val Ala Ser Gln Thr Pro Asp Leu Lys Ala Ser Ala Cys Leu Gly  
 20 25 30

Leu Pro Lys Cys Trp Asp  
 35

<210> 201  
 <211> 53  
 <212> PRT  
 <213> Homo sapiens

<220>



&lt;221&gt; UNSURE

&lt;222&gt; (12)

&lt;220&gt;

&lt;221&gt; UNSURE

&lt;222&gt; (15)

&lt;400&gt; 201

Met Ser Phe Leu Phe Leu Asp Ile Ala Lys Trp Xaa Phe Phe Xaa Phe  
 1 5 10 15

Leu Phe Cys Tyr Cys Phe Leu Ile Tyr Tyr Lys Met Leu Phe Phe Tyr  
 20 25 30

Gly Gly Phe Lys His Pro Ile Pro Cys Pro Gly Phe Leu His His Trp  
 35 40 45

Ile Leu Leu Ile Ile  
 50

&lt;210&gt; 202

&lt;211&gt; 59

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 202

Met Gln Leu Trp Gly Glu Tyr Ser Pro Tyr Phe Cys Arg Asn Asn Asn  
 1 5 10 15

Phe Glu Tyr Leu Cys Ala Thr Thr Val Ala Asn Thr Arg Leu Arg Cys  
 20 25 30

Leu Leu Leu Leu Ser Gln Pro Cys Glu Val Lys Thr Leu Ser Leu Leu  
 35 40 45

Thr Asp Glu Glu Thr Asp Ser Glu Asp Ile Lys  
 50 55

&lt;210&gt; 203

&lt;211&gt; 18

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 203

Met Arg Cys Thr Gln Gln Phe Ser Ile Leu Ala Val Phe Lys Cys Thr  
 1 5 10 15

Ile Gln

&lt;210&gt; 204

&lt;211&gt; 177

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 204

Met Asn Phe Leu Lys Leu Ile Ala Val Phe Ile Val Phe Ser His Ala  
 1 5 10 15

Ser Glu Ser Pro Gln Asp Ser Thr Pro Asn Gln Leu Tyr Ile Trp Gly  
                   20                  25                  30  
 Arg Thr Lys Ala Leu Val Phe Phe Arg Ser Ser Thr Gly Asp Ser Asp  
                   35                  40                  45  
 Ser Thr Ala Arg Ile Lys Lys Leu Ile Asn Gly Asn Ser Met Pro Val  
           50                  55                  60  
 Ala Glu Glu Leu Pro Trp Glu Met Ser His Thr Glu His Gln Ser Ser  
           65                  70                  75                  80  
 Phe Pro Thr Pro Glu Ile Pro His Ser Leu Ala Pro Gly Thr Val Ala  
                   85                  90                  95  
 Ile Ser Lys Pro Trp Phe Pro Ala Val Ser Gln Ile Ala Arg Val Gln  
                   100                  105                  110  
 Arg Val Asp Ile Asn Phe Cys Ser Trp Glu Asp Leu Ser Pro Ser Gly  
           115                  120                  125  
 Lys Ala Thr Gly Lys Ser Arg Thr His Cys Thr Val Thr Ala Val Ser  
           130                  135                  140  
 Ser Asn Ala Thr Thr His Ala Gly Ile Asn Asn Glu His Gly Trp Gly  
   145                  150                  155                  160  
 Ser Leu Glu Leu Leu Asn Cys Lys Ala His Lys Cys Leu Asn Phe Phe  
                   165                  170                  175  
 His

<210> 205  
 <211> 119  
 <212> PRT  
 <213> Homo sapiens

<400> 205  
 Met Thr Ser Met Ala Glu Pro Gly Leu Ala Leu Tyr Leu Cys Gly His  
   1                  5                  10                  15  
 Thr Val Val Trp Ser Ser Ser Ser Leu Met Val Thr Phe Val Arg Ile  
           20                  25                  30  
 Leu Ile Ser Val Phe Phe Leu Pro Gln Phe Ser Ser Ser Arg Leu Pro  
           35                  40                  45  
 His Pro Cys Ser Leu Phe Met Pro Ala Trp Val Val Ala Leu Asp Glu  
           50                  55                  60  
 Thr Ala Val Thr Val Gln Cys Val Leu Leu Phe Pro Val Ala Phe Pro  
           65                  70                  75                  80  
 Leu Gly Glu Arg Ser Ser His Glu Gln Lys Phe Ile Ser Thr Arg Trp  
           85                  90                  95  
 Thr Leu Ala Ile Cys Glu Thr Ala Gly Asn Gln Gly Leu Leu Ile Ala  
           100                  105                  110

Thr Val Pro Gly Ala Lys Glu  
115

<210> 206  
<211> 33  
<212> PRT  
<213> Homo sapiens

<400> 206  
Met Leu Ile Ser Lys Ile Ile Ile Gly Ile Lys Thr Gln Arg Tyr Leu  
1 5 10 15

Ile Glu Lys Ser His Arg Ser Pro Arg Ile Tyr Ile Tyr Leu Gly Leu  
20 25 30

Ala

<210> 207  
<211> 126  
<212> PRT  
<213> Homo sapiens

<400> 207  
Leu Pro Cys Ser Asn Phe Phe Phe Phe Ser Phe Ser Leu Phe Leu Val  
1 5 10 15

Phe Ile Phe Ser Ala Ile Ser Arg Ile Phe Leu Leu Leu Ala Met Ser  
20 25 30

Gln Ser Ile Met Ala Leu Ser Pro Arg Leu Glu Cys Asn Gly Ala Val  
35 40 45

Ser Gly His Cys Asn Pro Cys Leu Pro Gly Ser Ser Asp Ser Pro Pro  
50 55 60

Ser Ala Ser Gln Val Ala Gly Ile Thr Gly Thr Cys His His Ala Arg  
65 70 75 80

Leu Ile Phe Val Phe Leu Val Glu Met Gly Phe His His Val Gly Gln  
85 90 95

Ala Gly Leu Glu Leu Leu Thr Ser Gly Asp Leu Pro Thr Ser Ala Ser  
100 105 110

Gln Ser Ala Gly Ile Thr Gly Val Ser His Arg Ala Arg Pro  
115 120 125

<210> 208  
<211> 88  
<212> PRT  
<213> Homo sapiens

<400> 208  
Met Val Tyr Lys Leu Glu Trp His Ile Ala Phe Leu Arg Ile Leu Arg  
1 5 10 15

Gln Arg Pro Gly Phe Gly Ala Lys Ile Lys Gly Trp Met Ser His Leu  
                             20                            25                            30

Pro Trp Tyr Gly Asn Ala Ser Val Leu Thr Ser Ala Gln Ser Asn Leu  
                             35                            40                            45

Lys Leu Ile Ser Pro Ser Arg Phe Phe Leu Leu Phe Leu Ala Arg Glu  
                             50                            55                            60

Lys Ile Thr Ser Ala Phe Phe Phe Arg Arg Val Lys Lys Lys Glu His  
                             65                            70                            75                            80

His Ser Ile Ser Gln Asn Cys Ile  
                             85

<210> 209  
 <211> 52  
 <212> PRT  
 <213> Homo sapiens

<400> 209  
 Met Ser Leu His Cys Val Thr Asn Thr Asp Leu Val Ser Lys Trp Cys  
                             1                            5                            10                            15

Arg Arg Thr Gln Ala Thr Thr Arg Asn Glu Pro Ser Leu Cys Asp Gln  
                             20                            25                            30

Gly Gly Pro Gly Arg Gln Thr Pro Ala His Glu Gly Arg Thr Val Val  
                             35                            40                            45

Ala Met Thr Ser  
                             50

<210> 210  
 <211> 63  
 <212> PRT  
 <213> Homo sapiens

<400> 210  
 Met Arg Leu Pro Asp Asp Ser Cys Pro Ser Cys Ser Gly Leu Pro Ala  
                             1                            5                            10                            15

Glu Lys Ser Cys Thr His Arg Ala Leu Leu Gly Phe Leu Thr Cys Gly  
                             20                            25                            30

Ile His Asp Pro Val Thr Pro Leu Ser Ser Val Met Val His Tyr Asn  
                             35                            40                            45

Asn Arg Ser Pro Asp His Gly Asn Tyr Phe Ser Ser Ser Thr Leu  
                             50                            55                            60

<210> 211  
 <211> 104  
 <212> PRT  
 <213> Homo sapiens

<400> 211  
 Met Asp Phe Glu Phe Ile Phe Phe Pro Leu Lys Lys Gly Asn Pro Leu

1	5	10	15
Ile Ala Lys Ser His Leu Gln Ile Val Lys Gln Thr Ser Gln Ile Thr			
20	25	30	
Lys Cys Phe Leu Cys Lys Gln Lys Ile Cys Phe Ala Gly Lys Gly Ile			
35	40	45	
Leu Leu Leu Asn Thr Gly Thr Val Ser Val Ile Leu Arg Met Gly Thr			
50	55	60	
Val Pro Tyr Asn Leu Phe Leu Lys Tyr Leu Leu Leu Gly Leu Ser			
65	70	75	80
Gln Ala Pro Ile Phe Ser Val Val Met Lys Lys Asn Tyr Gln Ala Thr			
85	90	95	
Ser Trp Val Phe Phe Ser Leu Phe			
100			

<210> 212  
 <211> 57  
 <212> PRT  
 <213> Homo sapiens

<400> 212
Met Ile Glu Leu Leu Ser Pro Tyr Gln Leu Arg Glu Leu Phe Cys Ser
1 5 10 15
Leu Thr His Val Gly Arg Thr Val Arg Trp Ser Glu Gln Trp Asn Leu
20 25 30
Leu Val Ala Gln Val Leu Glu Val Tyr Ser Asn Gly Gly Arg Thr Gln
35 40 45
Leu Gly Ile Trp Phe Leu Leu Ser Lys
50 55

<210> 213  
 <211> 31  
 <212> PRT  
 <213> Homo sapiens

<400> 213
Met Leu Glu Phe Gly Lys Cys Lys Phe Cys Phe Ala Asp Glu Ile Phe
1 5 10 15
Leu Leu Asn Phe Asn Thr Leu Lys Gly Ile Pro Val Phe Asn Tyr
20 25 30

<210> 214  
 <211> 37  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> UNSURE  
 <222> (29) .. (32)

&lt;220&gt;

&lt;221&gt; UNSURE

&lt;222&gt; (36)

&lt;400&gt; 214

Met Leu Ile Glu Val Phe Lys Gly Ile Tyr Lys Leu Asn Thr Leu His  
 1 5 10 15

Asn Tyr Gln Leu Asn Lys Cys Phe Tyr His Met Gln Xaa Xaa Xaa Xaa  
 20 25 30

Phe Phe Leu Xaa Arg  
 35

&lt;210&gt; 215

&lt;211&gt; 131

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 215

Met Gly Gln Lys Ile Ser Arg Gln Pro Tyr Ser Gly Ser Trp Ser Leu  
 1 5 10 15

Phe Ser Cys Ser Asp Pro Gln Lys Ala Ser Lys Ser Leu Asn Leu Glu  
 20 25 30

Thr Arg Gln Phe Phe Leu Ile Ser Cys Leu Lys Ala Val Gln Ser Ser  
 35 40 45

Val Asn Lys Pro Leu His Ala Gly Leu Ile Asn Ala Gly Pro Leu Arg  
 50 55 60

Ala Met Thr Gln Glu His Gly Leu Gly Ser Thr Leu Lys Ser Arg Asn  
 65 70 75 80

His Ser Thr Asp Asn Gly Asn Phe Val Gly Gly Asn Arg Leu Leu Glu  
 85 90 95

Leu Asn Ala Phe Val Arg Phe Leu Asp Leu Gln Ile Ser Leu Cys Gly  
 100 105 110

Pro Ala Leu Gly Gly Lys Ala Gly Ile His Asn Asn Leu Ile Asn Leu  
 115 120 125

Thr Gln Thr  
 130

&lt;210&gt; 216

&lt;211&gt; 57

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 216

Met Glu Phe Arg Cys Gln Leu Ile Pro Arg Leu Ile Leu Ser Tyr Ile  
 1 5 10 15

Lys Val Asn Asp Ile Leu His Glu Ile Met Leu Val Glu Pro Thr Arg  
 20 25 30

Leu Leu Ala Met Leu Pro Ser Leu Ser Ser Leu Asp Phe Leu Phe Lys  
                   35                  40                  45

Ser Leu Tyr Arg Val Thr Val Glu His  
           50                  55

<210> 217  
 <211> 67  
 <212> PRT  
 <213> Homo sapiens

<400> 217  
 Met Cys Glu Leu Pro Leu Leu Leu Cys Asn Ser Ile Leu Phe Met Ile  
       1                  5                  10                  15

Cys Asp Val Ile Arg Lys Phe Leu Leu Met Cys Gln Asn Lys Phe Asn  
                   20                  25                  30

Phe Pro Leu Arg Gln Phe Ile Thr Leu Phe Lys Trp Asn Ile Lys Glu  
                   35                  40                  45

Glu Pro Pro Ile Cys Lys Ile Leu Thr Phe Lys Phe Met Leu Ile Phe  
           50                  55                  60

Leu Asn Tyr  
       65

<210> 218  
 <211> 69  
 <212> PRT  
 <213> Homo sapiens

<400> 218  
 Met Ser Cys Leu Ser Tyr Gly Phe Lys Tyr Leu Gln Cys Ile Ala Lys  
       1                  5                  10                  15

Tyr Cys Ser Cys Thr Leu Gln Leu Arg Asn Thr Val Leu Gly Phe Gln  
                   20                  25                  30

Gln Lys Tyr Leu Arg Ile Ser His Ser Ser Leu Lys Lys Asp Ala Lys  
                   35                  40                  45

Asp Val Thr Gly Ile Ile Ile Val Ala Val Ser Cys Arg Ile Lys Asp  
           50                  55                  60

Arg Thr Arg Tyr Gly  
       65

<210> 219  
 <211> 29  
 <212> PRT  
 <213> Homo sapiens

<400> 219  
 Met Leu Trp Ser Leu Tyr Ile Ser Phe Lys Val Val Ala Asn Lys Arg  
       1                  5                  10                  15

Met Pro Ile Gln Gly Ile Tyr Trp His Phe His Gly Gly  
                             20                            25

<210> 220  
 <211> 26  
 <212> PRT  
 <213> Homo sapiens

<400> 220  
 Met Asn Phe Asp Cys Ala Ser Ala Ile Leu Asp Ile Phe Val Met Ile  
           1                            5                            10                            15

Gly Asn Arg Thr Ile Lys Cys Leu Ala Leu  
                             20                            25

<210> 221  
 <211> 41  
 <212> PRT  
 <213> Homo sapiens

<400> 221  
 Met Leu Phe Leu Asn Trp Ala Pro Ser Ser Asp Phe Ala Asn Leu Lys  
           1                            5                            10                            15

Ser Ile Thr Cys Leu Cys Leu Ser Lys Asn Pro Ser Ile Pro Ser Ser  
                             20                            25                            30

Leu Ile Ala Pro Cys Tyr Ser Pro Val  
                             35                            40

<210> 222  
 <211> 45  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> UNSURE  
 <222> (31)..(39)

<220>  
 <221> UNSURE  
 <222> (42)..(43)

<400> 222  
 Met Thr Ile Trp Gln Arg Tyr Phe Ser Tyr Asn Glu Lys Tyr Leu Cys  
           1                            5                            10                            15

Pro Ile Ser Leu Lys Ser Asp Val Glu Lys Leu Tyr Ile Tyr Xaa Xaa  
                             20                            25                            30

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Ile Leu Xaa Xaa Leu Leu  
                             35                            40                            45

<210> 223  
 <211> 31  
 <212> PRT  
 <213> Homo sapiens



&lt;400&gt; 223

Met Phe Gln Ser Val Arg Glu Met Ser Leu Ser Gly Ser Ile Pro Ala  
 1 5 10 15

Asn Asn Glu Glu Gly Met Arg Gln Ala Gln Trp His Ser Arg Leu  
 20 25 30

&lt;210&gt; 224

&lt;211&gt; 48

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 224

Phe Phe Phe Phe Phe Leu Arg Gln Ser Phe Thr Leu Ser Gln Ala Gly  
 1 5 10 15

Val Ala Trp His Asp Leu Gly Ser Leu His Pro Pro Leu Pro Gly Ser  
 20 25 30

Ser Asp Ser Arg Ala Ser Ala Ser Gln Ser Ala Arg Ile Thr Gly Val  
 35 40 45

&lt;210&gt; 225

&lt;211&gt; 30

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 225

Met Tyr Gln Lys Lys Pro Ile Arg Leu Lys Val Leu Lys Thr Arg Tyr  
 1 5 10 15

Lys Tyr Ser His Arg Tyr Val Ser Glu Thr Tyr Leu Phe Gln  
 20 25 30

&lt;210&gt; 226

&lt;211&gt; 44

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 226

Met Asn Gln Asn Leu His His Leu Tyr Asn Lys Arg Ser Glu Ser Ile  
 1 5 10 15

Ala Cys Leu Ala Trp His Val Gly Arg Val Ala Lys Asp Gln Cys Ser  
 20 25 30

Leu Met Tyr Phe Phe Lys Leu Ser Asn Asn Ile His  
 35 40

&lt;210&gt; 227

&lt;211&gt; 57

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 227

Met Leu Ile Ser Phe Trp Leu Leu Thr His Ala Ala Phe Ser Gly His  
 1 5 10 15

His Met Ala Leu Lys Gln Arg Ser Val Cys Ile His Ser Pro Tyr Glu  
 20 25 30

Ala Tyr Val Asn Ile Asn His Gly Met Phe Pro Asn Ile Leu Leu Ile  
 35 40 45

Phe Ala Ser Gln Leu Gly Ser Leu Ile  
 50 55

&lt;210&gt; 228

&lt;211&gt; 101

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; UNSURE

&lt;222&gt; (32)..(73)

&lt;400&gt; 228

Met Phe His Val Phe Ser Cys Ser Arg Ser Asp Leu Ala Thr Pro Gly  
 1 5 10 15

Asp Thr Phe Gly Tyr Thr Asn Arg Val Tyr Leu Gly Gln Arg Trp Xaa  
 20 25 30

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
 35 40 45

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
 50 55 60

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Leu Asp Ser Pro Arg Trp Ile  
 65 70 75 80

Ser Pro Leu Ser Pro Thr Met Leu Val Leu Leu Thr Trp Leu Leu Ile  
 85 90 95

Lys Gln Cys Gln Val  
 100

&lt;210&gt; 229

&lt;211&gt; 88

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 229

Met Leu His Ser Pro Gly Leu Thr Arg Gly Trp Pro Gln Lys Arg Val  
 1 5 10 15

Gly Glu Ala Gly Gln Gln Gly Leu Ala Glu Ile Ile Cys Arg Ala Gln  
 20 25 30

Glu Ala Gly Glu Arg Arg Gln Phe Gln Gly Pro Phe Val Arg Gln Val  
 35 40 45

Pro Gly Ala Gln Pro Gly Arg Gln Glu Gly Leu Ser Pro Ser Pro Arg  
           50                          55                          60

Gln Glu Gly Ser Gln Ala Glu Ala Pro Pro Ser Gly Thr Pro Gln Pro  
       65                          70                          75                          80

Thr Pro Ala Ala Leu Gly Gln Asp  
                           85

<210> 230  
 <211> 23  
 <212> PRT  
 <213> Homo sapiens

<400> 230  
 Met Glu Ala Thr Gly Val Thr Phe Ser Ser Phe Val Phe Glu Gln Gly  
       1                          5                          10                          15

Met Ser Val Leu Ser Leu Lys  
                           20

<210> 231  
 <211> 48  
 <212> PRT  
 <213> Homo sapiens

<400> 231  
 Met Lys Thr Glu Asp Ile Lys Cys Ala Arg Val Arg Ser Leu Ser His  
       1                          5                          10                          15

Ala Lys Gly Lys Val Lys Ile Ala Phe Phe His Ile Val Ser Glu Val  
                           20                          25                          30

Gln Leu Leu Arg Leu Ile Asn Glu Ser Cys Ser Ile Lys Gly Leu Thr  
           35                          40                          45

<210> 232  
 <211> 25  
 <212> PRT  
 <213> Homo sapiens

<400> 232  
 Met Arg Tyr Ile His Val Glu Phe Cys Ser Cys Gly Leu Met Ile Phe  
       1                          5                          10                          15

Thr Leu Tyr Ser Ile Thr Phe His Gly  
           20                          25

<210> 233  
 <211> 55  
 <212> PRT  
 <213> Homo sapiens

&lt;400&gt; 233

Met Leu Pro Thr Pro Val Pro Thr Ile Glu Ala Leu Leu Phe Met Leu  
 1 5 10 15

Lys Cys Gln Val Leu Thr Val His Gly Ser Met Glu Thr Phe Leu Leu  
 20 25 30

Phe Ser Val Val Leu Gly Ala Ser Leu Leu Val Asn Leu Arg Lys Ile  
 35 40 45

Gly Asp Ser Val Asn Leu Glu  
 50 55

&lt;210&gt; 234

&lt;211&gt; 148

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 234

Met Gly Arg Ile Arg Pro Asp His Thr Leu Leu Phe Gln Arg Gly Pro  
 1 5 10 15

Val Pro Ala Pro Leu Thr Ser Gly Leu His Tyr Tyr Thr Thr Leu Glu  
 20 25 30

Glu Leu Trp Lys Ser Phe Asp Leu Cys Glu Asp Tyr Phe Lys Pro Pro  
 35 40 45

Phe Gly Pro Tyr Pro Glu Lys Ser Gly Lys Asp Ser Leu Val Ser Met  
 50 55 60

Lys Cys Ser Leu Phe Arg Phe Cys Pro Trp Ser Lys Glu Leu Pro Phe  
 65 70 75 80

Gln Pro Pro Glu Gly Ser Ile Ser Ser His Leu Gly Ser Gly Ala Ser  
 85 90 95

Asp Ser Glu Thr Glu Glu Thr Arg Lys Ala Leu Pro Ile Gln Ser Phe  
 100 105 110

Ser His Glu Lys Glu Ser His Gln His Arg Gln His Ser Val Pro Val  
 115 120 125

Ile Ser Arg Pro Gly Ser Asn Val Lys Pro Thr Leu Pro Pro Ile Pro  
 130 135 140

Gln Gly Arg Arg  
 145

&lt;210&gt; 235

&lt;211&gt; 940

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 235

Glu Tyr Thr Ser Phe Ser Ala Leu His Asn Thr Tyr Ser Lys Ile Asp  
 1 5 10 15

His Ile Val Gly Ser Lys Ala Leu Leu Ser Lys Cys Lys Arg Thr Glu

	<b>20</b>						<b>25</b>						<b>30</b>						
Met	Ile	Thr	Asn	Cys	Leu	Ser	Asp	His	Ser	Ala	Ile	Lys	Leu	Glu	Leu				
	35						40					45							
Arg	Ile	Lys	Lys	Leu	Thr	Gln	Asn	Cys	Ser	Thr	Thr	Trp	Lys	Leu	Asn				
	50					55					60								
Asn	Leu	Leu	Leu	Asn	Asp	Tyr	Cys	Val	His	Asn	Lys	Met	Lys	Ala	Glu				
	65				70					75					80				
Ile	Lys	Met	Phe	Phe	Glu	Thr	Asn	Glu	Asn	Lys	Asp	Thr	Thr	Tyr	Gln				
			85						90					95					
Asn	Leu	Trp	Asp	Thr	Phe	Lys	Ala	Val	Cys	Arg	Gly	Asn	Phe	Ile	Ala				
			100					105					110						
Leu	Asn	Val	His	Lys	Arg	Lys	Gln	Glu	Arg	Ser	Lys	Ile	Asp	Thr	Leu				
		115					120					125							
Ile	Ser	Gln	Leu	Lys	Glu	Leu	Glu	Lys	Gln	Glu	Gln	Thr	His	Ser	Lys				
	130					135					140								
Ala	Ser	Arg	Arg	Gln	Glu	Ile	Thr	Lys	Ile	Arg	Ala	Glu	Val	Lys	Glu				
	145				150					155					160				
Ile	Glu	Thr	Gln	Lys	Thr	Phe	Lys	Arg	Ile	Asn	Glu	Ser	Arg	Asn	Trp				
			165					170						175					
Phe	Phe	Glu	Arg	Ile	Ser	Lys	Ile	Asp	Arg	Pro	Leu	Ala	Arg	Leu	Ile				
			180					185					190						
Lys	Lys	Lys	Arg	Glu	Lys	Asn	Gln	Ile	Asp	Ala	Ile	Asn	Thr	His	Asp				
		195					200					205							
Lys	Gly	Asp	Ile	Thr	Thr	Asp	Pro	Thr	Glu	Ile	Gln	Thr	Thr	Ile	Arg				
	210					215					220								
Glu	Tyr	Tyr	Lys	His	Phe	Tyr	Ala	Asn	Lys	Leu	Glu	Asn	Leu	Glu	Glu				
	225				230					235					240				
Met	Asp	Lys	Phe	Leu	Asp	Thr	Tyr	Thr	Leu	Pro	Arg	Leu	Asn	Gln	Glu				
				245					250					255					
Glu	Ala	Glu	Ser	Leu	Asn	Arg	Pro	Ile	Thr	Asp	Ser	Glu	Ile	Ala	Ala				
			260					265					270						
Ile	Ile	Asn	Ser	Leu	Pro	Thr	Lys	Lys	Ser	Pro	Gly	Pro	Asp	Gly	Phe				
		275					280					285							
Thr	Pro	Lys	Phe	Tyr	Gln	Arg	Tyr	Lys	Glu	Glu	Leu	Val	Pro	Phe	Leu				
	290					295					300								
Leu	Lys	Leu	Phe	Gln	Ser	Ile	Thr	Lys	Glu	Gly	Ile	Leu	Pro	Asn	Ser				
				310						315					320				
Phe	Tyr	Glu	Ala	Asn	Ile	Ile	Leu	Ile	Leu	Lys	Pro	Gly	Arg	Asp	Thr				
				325					330					335					
Thr	Lys	Lys	Arg	Glu	Phe	Arg	Pro	Ile	Ser	Met	Met	Ile	Ile	Asp	Ala				
			340					345					350						

Lys Ile Leu Ser Lys Ile Leu Ala Asn Gln Ile Gln Gln His Leu Ile  
 355 360 365  
 Lys Leu Ile His His Asp Gln Val Gly Phe Ile Pro Gly Met Lys Gly  
 370 375 380  
 Trp Phe Asn Ile Arg Lys Ser Ile Lys Val Ile His His Ile Asn Arg  
 385 390 395 400  
 Thr Lys Asp Lys Asn His Met Ile Ile Ser Ile His Ala Glu Lys Ala  
 405 410 415  
 Phe Asp Lys Ile Gln Gln Pro Phe Met Leu Lys Thr Val Asn Lys Leu  
 420 425 430  
 Val Ile Asp Gly Thr Tyr Leu Lys Ile Ile Arg Ala Ile Tyr Asp Lys  
 435 440 445  
 Pro Thr Ala Asn Ile Ile Leu Asn Gly Gln Lys Leu Glu Ala Phe Pro  
 450 455 460  
 Leu Arg Thr Gly Ile Arg Gln Gly Cys Pro Leu Ser Pro Leu Leu Phe  
 465 470 475 480  
 Asn Ile Val Leu Glu Val Leu Ala Arg Ala Ile Arg Gln Glu Lys Glu  
 485 490 495  
 Ile Lys Gly Ile Gln Leu Gly Lys Glu Lys Val Lys Leu Ser Leu Phe  
 500 505 510  
 Ala Asp Asp Met Ile Leu Tyr Leu Glu Asn Pro Ile Val Ser Ala Gln  
 515 520 525  
 Asn Leu Leu Lys Leu Met Ser Ser Phe Ser Lys Val Ser Gly Tyr Lys  
 530 535 540  
 Ile Asn Val Gln Lys Ser Gln Ala Phe Leu Tyr Thr Asn Asn Arg Gln  
 545 550 555 560  
 Thr Glu Ser Gln Met Ser Glu Leu Pro Phe Ala Ile Ala Ser Lys Arg  
 565 570 575  
 Ile Lys Tyr Leu Gly Ile Gln Leu Thr Arg Asp Val Lys Asp Leu Phe  
 580 585 590  
 Lys Glu Asn Tyr Lys Pro Leu Leu Asn Lys Ile Lys Glu Asp Thr Asn  
 595 600 605  
 Lys Trp Lys Asn Ile Pro Cys Ser Trp Ile Gly Arg Ile Asn Ile Val  
 610 615 620  
 Lys Met Ala Ile Met Pro Lys Val Ile Tyr Arg Phe Asn Ala Ile Pro  
 625 630 635 640  
 Ile Lys Leu Pro Met Thr Phe Phe Thr Glu Leu Glu Lys Thr Thr Leu  
 645 650 655  
 Lys Phe Ile Trp Asn Gln Lys Arg Ala Arg Ile Ala Lys Thr Ile Leu  
 660 665 670

Ser Gln Lys Asn Lys Ala Gly Gly Ile Thr Leu Pro Asp Phe Lys Leu  
 675 680 685  
 Tyr Tyr Lys Ala Thr Val Thr Lys Thr Ala Trp Tyr Trp Tyr Gln Asn  
 690 695 700  
 Arg Asp Ile Asp Gln Trp Asn Arg Ile Glu Pro Leu Glu Leu Ile Pro  
 705 710 715 720  
 His Ile Tyr Asn His Leu Ile Phe Asp Lys Pro Asp Lys Asn Lys Leu  
 725 730 735  
 Trp Gly Lys Asp Ser Leu Phe Asn Lys Trp Cys Trp Glu Asn Trp Leu  
 740 745 750  
 Ala Ile Cys Arg Lys Leu Lys Leu Asn Leu Phe Leu Thr Pro Tyr Thr  
 755 760 765  
 Lys Ile Asn Ser Arg Trp Ile Lys Asp Leu Asn Val Arg Pro Lys Thr  
 770 775 780  
 Ile Lys Ile Leu Glu Lys Asn Leu Gly Asn Thr Ile Gln Asp Ile Gly  
 785 790 795 800  
 Val Gly Lys Asp Phe Met Thr Lys Thr Pro Lys Ala Met Ala Thr Lys  
 805 810 815  
 Ala Lys Ile Asp Lys Trp Asp Ile Ile Lys Leu Lys Ser Phe Cys Thr  
 820 825 830  
 Ala Lys Glu Thr Thr Ile Ile Val Asn Arg Gln Pro Thr Glu Trp Glu  
 835 840 845  
 Lys Ile Phe Lys Ile Tyr Pro Ser Asp Lys Gly Leu Ile Ser Arg Ile  
 850 855 860  
 Tyr Lys Glu Leu Lys Gln Ile Tyr Lys Lys Lys Ser Asn Asn Pro Ile  
 865 870 875 880  
 Lys Asn Trp Ala Lys Asp Met Asn Arg His Phe Ser Lys Glu Asp Ile  
 885 890 895  
 Tyr Ala Val Asn Arg His Met Lys Thr Cys Ser Ser Leu Leu Ala Ile  
 900 905 910  
 Arg Glu Met Gln Ile Lys Thr Thr Met Arg Tyr His Phe Thr Pro Val  
 915 920 925  
 Arg Met Ala Ser Ile Lys Lys Ser Gly Asn Asn Arg  
 930 935 940

&lt;210&gt; 236

&lt;211&gt; 58

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 236

Met Ala Ile Glu Val Cys Trp Pro Leu Pro Leu Asp Gly Leu Leu Leu  
 1 5 10 15

Leu Ala Leu Glu Phe Leu Arg Pro Leu Phe Ile Ile Pro Gln Ser Phe  
                     20                    25                    30  
 Phe Leu Leu Pro Ala Met Leu Cys Leu Phe Phe Ala Leu Leu Ser Pro  
                     35                    40                    45  
 Arg Thr Thr Phe Phe His Phe His Ser Gly  
                     50                    55

<210> 237  
 <211> 34  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> UNSURE  
 <222> (28)

<400> 237  
 Met Pro Leu His Leu Gly Tyr Lys Val Ser Pro Pro Pro Gln Ser His  
                     1                    5                    10                    15  
 Gly Leu Ala Asn Tyr Leu Ser Val Phe Asp Cys Xaa Val Val Ser Thr  
                     20                    25                    30  
 Gly Glu

<210> 238  
 <211> 27  
 <212> PRT  
 <213> Homo sapiens

<400> 238  
 Met Arg Lys Val Cys Val Pro Ala Phe Met Thr Ile Glu Ser Arg Gln  
                     1                    5                    10                    15  
 Leu Leu Ser Gly Val Ser Ala Cys Phe Gln Gln  
                     20                    25

<210> 239  
 <211> 26  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> UNSURE  
 <222> (22)

<400> 239  
 Met Thr Ser Ile Thr Val Leu Phe Ser Lys Lys Arg Leu Ser Leu Met  
                     1                    5                    10                    15  
 Ala Ser Arg Cys Val Xaa Leu Met Arg Tyr  
                     20                    25

<210> 240



<211> 45  
 <212> PRT  
 <213> Homo sapiens

<400> 240  
 Met Lys Ser Gln Leu Gln Ser Leu His Pro Phe Phe Ser Lys Leu Ala  
           1                          5                          10                          15  
 Leu Leu Val Ser Val Leu Phe Tyr Ile Ile Trp Leu His Leu Thr Val  
                           20                          25                          30  
 Phe Lys Lys Ser Ser Val Leu Gln Lys Asn Phe Lys Leu  
                           35                          40                          45

<210> 241  
 <211> 65  
 <212> PRT  
 <213> Homo sapiens

<400> 241  
 Met Ile Gly Ile Thr Trp Cys Phe Glu Leu Ile His Pro Thr Leu Glu  
           1                          5                          10                          15  
 Leu Thr Ala Thr Val Pro Asp Phe His Arg Tyr Ala Ser Phe His Ser  
                           20                          25                          30  
 Gly Ser Leu Pro Glu Val Leu His Ser Gly Glu His Ala Gln Val Ser  
                           35                          40                          45  
 Pro Ala Leu Gln Asn His Pro Glu Cys Gln Arg Leu Gln His Lys Gly  
           50                          55                          60  
 Lys  
   65

<210> 242  
 <211> 42  
 <212> PRT  
 <213> Homo sapiens.

<400> 242  
 Ile Phe Thr Ala Met Pro Pro Phe Thr Leu Gly Val Phe Gln Arg Ser  
           1                          5                          10                          15  
 Cys Thr Arg Glu Ser Met Leu Arg Phe Pro Gln Leu Tyr Lys Ile Thr  
                           20                          25                          30  
 Gln Asn Ala Lys Asp Phe Asn Thr Arg Val  
           35                          40

<210> 243  
 <211> 40  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> UNSURE  
 <222> (2)

&lt;400&gt; 243

Met Xaa Leu Val Leu Leu Thr Arg Leu Ile Arg Arg Ser Leu Tyr Thr  
 1 5 10 15

Lys Arg Asn Leu Leu Ser His Ser His Asn Lys Thr Ser His Gln Thr  
 20 25 30

Asn Asp Thr Lys Ser Glu Asn His  
 35 40

&lt;210&gt; 244

&lt;211&gt; 56

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 244

Met Phe Pro Glu Leu Ala Ser Leu Tyr Pro Gly Lys Gly Thr Ser Phe  
 1 5 10 15

Ser Trp Ala Val Pro Pro Pro Gln Lys Pro Glu Ser Gln Pro Cys Arg  
 20 25 30

Val Pro Ser Ser Ser Phe Gln Ile Gln Ile Thr Pro Thr Ser Ser Leu  
 35 40 45

Gly Ser Pro Ser Leu Arg Thr Gln  
 50 55

&lt;210&gt; 245

&lt;211&gt; 26

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 245

Met Lys Lys Pro Glu Ala Glu Ala Ala Leu Thr Leu Arg Asn Pro Val  
 1 5 10 15

Ser Gln Arg Asp Leu Ala Ile Leu Ala Ser  
 20 25

&lt;210&gt; 246

&lt;211&gt; 43

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 246

Met Pro Ile Tyr Pro Cys Pro Cys Arg Val Gly Arg Lys Asn Leu Met  
 1 5 10 15

Leu Ala Asn Ser Pro His Phe Asn Ser Thr Leu Gln Thr Leu Ser Lys  
 20 25 30

Cys Leu Leu Phe Val Arg Gln Tyr Ala Ser His  
 35 40

&lt;210&gt; 247

<211> 49  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> UNSURE  
 <222> (12)..(33)

<400> 247  
 Met Lys Gln Trp Asp Ala Val Arg Lys Arg Lys Xaa Xaa Xaa Xaa Xaa  
     1                    5                    10                    15  
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
                     20                    25                    30  
 Xaa Cys Arg Gly Lys Val Asn Lys Asn Cys Ile Ile Leu Gly Val Phe  
                     35                    40                    45

Cys

<210> 248  
 <211> 24  
 <212> PRT  
 <213> Homo sapiens

<400> 248  
 Met Pro Tyr Asp Ser Thr Tyr Ile Lys Ser Lys His Gln Ala Val Leu  
     1                    5                    10                    15  
 Ser Met Ile Val Lys Leu Val Gly  
                     20

<210> 249  
 <211> 30  
 <212> PRT  
 <213> Homo sapiens

<400> 249  
 Met His Ile Ser Phe Gly Ile Gln Ile Ile Val Asn Asp Gly Glu Leu  
     1                    5                    10                    15  
 Thr Ser Asn Ile Ser Ser Tyr Thr Thr Asn Val Ile Lys Pro  
                     20                    25                    30

<210> 250  
 <211> 192  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> UNSURE  
 <222> (136)

<400> 250  
 Met Pro Ser Val Arg Ala His Pro Asn Pro Arg Ala Glu Gly His Glu  
     1                    5                    10                    15

Gly Ala Lys Ser Leu Arg Asn Ala Ile Leu Arg Leu Val Arg Asp Met  
                   20                  25                  30  
 Glu Ile Arg Thr Gln Gly Gly Pro Gly Leu Gly Asn Asp Trp Glu Thr  
                   35                  40                  45  
 Cys Leu Gly Ser Gln Asp Leu Gly Val Leu Thr Pro Ser Pro His Pro  
                   50                  55                  60  
 Ala Val Pro Ser Val Pro Ser Pro Ser Leu Ser Lys Pro Leu Gly Ile  
                   65                  70                  75                  80  
 Glu Trp Pro Leu Leu Phe Trp Cys Pro Gly Val Ile Val Pro Lys Leu  
                   85                  90                  95  
 Leu Phe Pro Val Pro Ser Pro Gln Arg Leu Val Arg Val Gly Met Arg  
                   100                  105                  110  
 Asp Gly Glu Gly Leu Gly Leu Trp Glu Gln Val Gly Gly Leu Ile Cys  
                   115                  120                  125  
 Gly Leu Ser Asp Ser Gln Leu Xaa Pro Arg Trp Gly Met Ser Pro Ser  
                   130                  135                  140  
 Leu Leu Ser Val Trp Val Arg Lys Thr Gly Cys Asp Pro Glu Glu Gly  
                   145                  150                  155                  160  
 Lys Ile Glu Lys Glu Gly Lys Asp Val Gly Glu Gly Gly Glu Arg Gln  
                   165                  170                  175  
 Asp Arg Arg Lys Glu Val Glu Glu Glu Val Val Gly Ile Gly Met Arg  
                   180                  185                  190

<210> 251  
 <211> 45  
 <212> PRT  
 <213> Homo sapiens

<400> 251  
 Met Gln Phe Cys Lys Ile Lys Cys Leu Ser Arg His Ala Tyr Asn Pro  
                   1                  5                  10                  15  
 Ala Ile Ala Cys Leu Gly Ala Tyr Leu Thr Glu Met Asn Ile Tyr Asn  
                   20                  25                  30  
 Tyr Ile Ile Ile Cys Thr Pro Asn Ser Ser Gln Leu Tyr  
                   35                  40                  45

<210> 252  
 <211> 169  
 <212> PRT  
 <213> Homo sapiens

<400> 252  
 Met Ala Pro Ser Glu Asp Pro Arg Asp Trp Arg Ala Asn Leu Lys Gly  
                   1                  5                  10                  15

Thr Ile Arg Glu Thr Gly Leu Glu Thr Ser Ser Gly Gly Lys Leu Ala  
                   20                  25                  30  
 Gly His Gln Lys Thr Val Pro Thr Ala His Leu Thr Phe Val Ile Asp  
                   35                  40                  45  
 Cys Thr His Gly Lys Gln Leu Ser Leu Ala Ala Thr Ala Ser Pro Pro  
           50                  55                  60  
 Gln Ala Pro Ser Pro Asn Arg Gly Leu Val Thr Pro Pro Met Lys Thr  
   65                  70                  75                  80  
 Tyr Ile Val Phe Cys Gly Glu Asn Trp Pro His Leu Thr Arg Val Thr  
                   85                  90                  95  
 Pro Met Gly Gly Gly Cys Leu Ala Gln Ala Arg Ala Thr Leu Pro Leu  
                  100                 105                 110  
 Cys Arg Gly Ser Val Ala Ser Ala Ser Phe Pro Val Ser Pro Leu Cys  
          115                 120                 125  
 Pro Gln Glu Val Pro Glu Ala Lys Gly Lys Pro Val Lys Ala Ala Pro  
   130                 135                 140  
 Val Arg Ser Ser Thr Trp Gly Thr Val Lys Asp Ser Leu Lys Ala Leu  
 145                 150                 155                 160  
 Ser Ser Cys Val Cys Gly Gln Ala Asp  
                  165

<210> 253  
 <211> 69  
 <212> PRT  
 <213> Homo sapiens

<400> 253  
 Met Phe Asn Val Arg Leu His Gln Asn Met Cys Gln Leu Thr Met Phe  
   1                  5                  10                  15  
 Asn Met Phe His Leu Gln Asn Phe Leu Glu Gly Lys Lys Ser Phe Leu  
          20                 25                 30  
 Val Asn Met Phe Phe Cys Leu Cys Phe Ile Ile Leu Ser Thr Met Asp  
          35                 40                 45  
 Thr Gly Asn Gln Ser Thr Val Asn Asn His Arg His His Phe Val Val  
   50                 55                 60  
 Leu Phe Leu Arg Val  
   65

<210> 254  
 <211> 33  
 <212> PRT  
 <213> Homo sapiens

<400> 254  
 Met Glu Val Arg Ser Val Ile Pro Gln Val Leu Asn Ala Trp Ala Ser

1                      5                      10                      15  
 Leu Met Ser Phe Tyr Gln Leu Ser Ala Thr Cys Val Lys Phe His Leu  
                     20                      25                      30  
 Ser

<210> 255  
 <211> 72  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> UNSURE  
 <222> (59)

<220>  
 <221> UNSURE  
 <222> (65)

<400> 255  
 Met Trp Thr Thr Cys Asn Val Thr Lys Gln Lys Glu Thr Gln Glu Ala  
   1                      5                      10                      15  
 Asn Ile Pro Ile Tyr Ser Pro Leu Ser Ala Leu Thr Gln Gln Asn Lys  
                     20                      25                      30  
 Thr Lys Pro Ala Thr Thr Ile Arg Phe Val Lys Ile Leu Val Val Arg  
                     35                      40                      45  
 Ile Pro Thr Leu Ser Ser Gln Gln Phe Gly Xaa Gln Lys Ser Leu Val  
                     50                      55                      60  
 Xaa Met Ser Val His Val Lys Ser  
   65                      70

<210> 256  
 <211> 131  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> UNSURE  
 <222> (31)..(93)

<220>  
 <221> UNSURE  
 <222> (111)..(121)

<400> 256  
 Met Tyr Ala Ser Asn Asn Leu Ser Arg Gly Arg Ile Pro Lys Glu Asn  
   1                      5                      10                      15  
 Ile Cys Ser Ser Phe Phe Leu Leu Arg Phe Phe Cys Ile Phe Xaa Xaa  
                     20                      25                      30  
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
                     35                      40                      45

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
 50 55 60

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
 65 70 75 80

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Thr Val Phe  
 85 90 95

Pro Leu Leu Ser Tyr Asn Asn Gln His Arg Arg Leu Leu Trp Xaa Gln  
 100 105 110

Met Trp Gly Asn Phe Phe His Ala Lys Xaa Ala Val Arg Ala Ala Val  
 115 120 125

Ser Pro Thr  
 130

<210> 257  
 <211> 44  
 <212> PRT  
 <213> Homo sapiens

<400> 257  
 Glu Ser Phe Tyr Asp Thr Phe His Thr Val Ala Asp Met Met Tyr Phe  
 1 5 10 15

Cys Gln Met Leu Ala Val Val Glu Thr Ile Asn Ala Ala Ile Gly Val  
 20 25 30

Thr Thr Ser Pro Val Leu Pro Ser Leu Ile Gln Val  
 35 40

<210> 258  
 <211> 70  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> UNSURE  
 <222> (8) .. (52)

<220>  
 <221> UNSURE  
 <222> (57)

<400> 258  
 Met Phe Ile Phe Thr Phe His Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
 1 5 10 15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
 20 25 30

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
 35 40 45

Xaa Xaa Xaa Xaa Cys Phe Phe Pro Xaa Trp Phe Leu Leu Phe Leu Leu  
 50 55 60

Arg Ser Val Ser Phe Cys  
65 70

<210> 259  
<211> 61  
<212> PRT  
<213> Homo sapiens

<220>  
<221> UNSURE  
<222> (16)..(53)

<400> 259  
Met Lys Ile Thr Tyr Leu Asp Ile Leu Glu Lys Tyr Ile His Ser Xaa  
1 5 10 15  
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
20 25 30  
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
35 40 45  
Xaa Xaa Xaa Xaa Xaa Glu Ser Thr Gln Ile Gly Pro Glu  
50 55 60

<210> 260  
<211> 2383  
<212> PRT  
<213> Homo sapiens

<400> 260  
Met Glu Thr Arg Ser Pro Gly Leu Asn Asn Met Lys Pro Gln Ser Leu  
1 5 10 15  
Gln Leu Val Leu Glu Glu Gln Val Leu Ala Leu Gln Gln Gln Met Ala  
20 25 30  
Glu Asn Gln Ala Ala Ser Trp Arg Lys Leu Lys Asn Ser Gln Glu Ala  
35 40 45  
Gln Gln Arg Gln Ala Thr Leu Val Arg Lys Leu Gln Ala Lys Val Leu  
50 55 60  
Gln Tyr Arg Ser Trp Cys Gln Glu Leu Glu Lys Arg Leu Glu Ala Thr  
65 70 75 80  
Gly Gly Pro Ile Pro Gln Arg Trp Glu Asn Val Glu Glu Pro Asn Leu  
85 90 95  
Asp Glu Leu Leu Val Arg Leu Glu Glu Glu Gln Gln Arg Cys Glu Ser  
100 105 110  
Leu Ala Gln Val Asn Thr Gln Leu Arg Leu His Met Glu Lys Ala Asp  
115 120 125  
Val Val Asn Lys Ala Leu Arg Glu Asp Val Glu Lys Leu Thr Val Asp  
130 135 140



Trp Ser Arg Ala Arg Asp Glu Leu Met Arg Lys Glu Ser Gln Trp Gln  
 145 150 155 160  
 Met Glu Gln Glu Trp Ser Leu Leu Phe Ser Leu Leu Val Leu Arg Asp  
 165 170 175  
 Leu Met Glu Leu Lys Ala Glu His Val Arg Leu Ser Gly Ser Leu Leu  
 180 185 190  
 Thr Cys Cys Leu Arg Leu Thr Val Gly Ala Gln Ser Arg Glu Pro Asn  
 195 200 205  
 Gly Ser Gly Arg Met Asn Gly Arg Glu Pro Ala Gln Leu Leu Leu Leu  
 210 215 220  
 Leu Ala Lys Thr Gln Glu Leu Glu Lys Glu Ala His Glu Arg Ser Gln  
 225 230 235 240  
 Glu Leu Ile Gln Leu Lys Ser Gln Gly Asp Leu Glu Lys Ala Glu Leu  
 245 250 255  
 Gln Asp Arg Val Thr Glu Leu Ser Ala Leu Leu Thr Gln Ser Gln Lys  
 260 265 270  
 Gln Asn Glu Asp Tyr Glu Lys Met Ile Lys Ala Leu Arg Glu Thr Val  
 275 280 285  
 Glu Ile Leu Glu Thr Asn His Thr Glu Leu Met Glu His Glu Ala Ser  
 290 295 300  
 Leu Ser Arg Asn Ala Gln Glu Glu Lys Leu Ser Leu Gln Gln Val Ile  
 305 310 315 320  
 Lys Asp Ile Thr Gln Val Met Val Glu Glu Gly Asp Asn Ile Ala Gln  
 325 330 335  
 Gly Ser Gly His Glu Asn Ser Leu Glu Leu Asp Ser Ser Ile Phe Ser  
 340 345 350  
 Gln Phe Asp Tyr Gln Asp Ala Asp Lys Ala Leu Thr Leu Val Arg Ser  
 355 360 365  
 Val Leu Thr Arg Arg Arg Gln Ala Val Gln Asp Leu Arg Gln Gln Leu  
 370 375 380  
 Ala Gly Cys Gln Glu Ala Val Asn Leu Leu Gln Gln His Asp Gln  
 385 390 395 400  
 Trp Glu Glu Glu Gly Lys Ala Leu Arg Gln Arg Leu Gln Lys Leu Thr  
 405 410 415  
 Gly Glu Arg Asp Thr Leu Ala Gly Gln Thr Val Asp Leu Gln Gly Glu  
 420 425 430  
 Val Asp Ser Leu Ser Lys Glu Arg Glu Leu Leu Gln Lys Ala Arg Glu  
 435 440 445  
 Glu Leu Arg Gln Gln Leu Glu Val Leu Glu Gln Glu Ala Trp Arg Leu  
 450 455 460  
 Arg Arg Val Asn Val Glu Leu Gln Leu Gln Gly Asp Ser Ala Gln Gly

465	470	475	480
Gln Lys Glu Glu Gln Gln Glu Glu Leu His Leu Ala Val Arg Glu Arg	485	490	495
Glu Arg Leu Gln Glu Met Leu Met Gly Leu Glu Ala Lys Gln Ser Glu	500	505	510
Ser Leu Ser Glu Leu Ile Thr Leu Arg Glu Ala Leu Glu Ser Ser His	515	520	525
Leu Glu Gly Glu Leu Leu Arg Gln Glu Gln Thr Glu Val Thr Ala Ala	530	535	540
Leu Ala Arg Ala Glu Gln Ser Ile Ala Glu Leu Ser Ser Ser Glu Asn	545	550	555
Thr Leu Lys Thr Glu Val Ala Asp Leu Arg Ala Ala Val Lys Leu	565	570	575
Ser Ala Leu Asn Glu Ala Leu Ala Leu Asp Lys Val Gly Leu Asn Gln	580	585	590
Gln Leu Leu Gln Leu Glu Glu Glu Asn Gln Ser Val Cys Ser Arg Met	595	600	605
Glu Ala Ala Glu Gln Ala Arg Asn Ala Leu Gln Val Asp Leu Ala Glu	610	615	620
Ala Glu Lys Arg Arg Glu Ala Leu Trp Glu Lys Asn Thr His Leu Glu	625	630	635
Ala Gln Leu Gln Lys Ala Glu Glu Ala Gly Ala Glu Leu Gln Ala Asp	645	650	655
Leu Arg Asp Ile Gln Glu Glu Lys Glu Glu Ile Gln Lys Lys Leu Ser	660	665	670
Glu Ser Arg His Gln Gln Glu Ala Ala Thr Thr Gln Leu Glu Gln Leu	675	680	685
His Gln Glu Ala Lys Arg Gln Glu Glu Val Leu Ala Arg Ala Val Gln	690	695	700
Glu Lys Glu Ala Leu Val Arg Glu Lys Ala Ala Leu Glu Val Arg Leu	705	710	715
Gln Ala Val Glu Arg Asp Arg Gln Asp Leu Ala Glu Gln Leu Gln Gly	725	730	735
Leu Ser Ser Ala Lys Glu Leu Leu Glu Ser Ser Leu Phe Glu Ala Gln	740	745	750
Gln Gln Asn Ser Val Ile Glu Val Thr Lys Gly Gln Leu Glu Val Gln	755	760	765
Ile Gln Thr Val Thr Gln Ala Lys Glu Val Ile Gln Gly Glu Val Arg	770	775	780
Cys Leu Lys Leu Glu Leu Asp Thr Glu Arg Ser Gln Ala Glu Gln Glu	785	790	795
			800

Arg Asp Ala Ala Ala Arg Gln Leu Ala Gln Ala Glu Gln Glu Gly Lys  
 805 810 815  
 Thr Ala Leu Glu Gln Gln Lys Ala Ala His Glu Lys Glu Val Asn Gln  
 820 825 830  
 Leu Arg Glu Lys Trp Glu Lys Glu Arg Ser Trp His Gln Gln Glu Leu  
 835 840 845  
 Ala Lys Ala Leu Glu Ser Leu Glu Arg Glu Lys Met Glu Leu Glu Met  
 850 855 860  
 Arg Leu Lys Glu Gln Gln Thr Glu Met Glu Ala Ile Gln Ala Gln Arg  
 865 870 875 880  
 Glu Glu Glu Arg Thr Gln Ala Glu Ser Ala Leu Cys Gln Met Gln Leu  
 885 890 895  
 Glu Thr Glu Lys Glu Arg Val Ser Leu Leu Glu Thr Leu Leu Gln Thr  
 900 905 910  
 Gln Lys Glu Leu Ala Asp Ala Ser Gln Gln Leu Glu Arg Leu Arg Gln  
 915 920 925  
 Asp Met Lys Val Gln Lys Leu Lys Glu Gln Glu Thr Thr Gly Ile Leu  
 930 935 940  
 Gln Thr Gln Leu Gln Glu Ala Gln Arg Glu Leu Lys Glu Ala Ala Arg  
 945 950 955 960  
 Gln His Arg Asp Asp Leu Ala Ala Leu Gln Glu Glu Ser Ser Ser Leu  
 965 970 975  
 Leu Gln Asp Lys Met Asp Leu Gln Lys Gln Val Glu Asp Leu Lys Ser  
 980 985 990  
 Gln Leu Val Ala Gln Asp Asp Ser Gln Arg Leu Val Glu Gln Glu Val  
 995 1000 1005  
 Gln Glu Lys Leu Arg Glu Thr Gln Glu Tyr Asn Arg Ile Gln Lys Glu  
 1010 1015 1020  
 Leu Glu Arg Glu Lys Ala Ser Leu Thr Leu Ser Leu Met Glu Lys Glu  
 1025 1030 1035 1040  
 Gln Arg Leu Leu Val Leu Gln Glu Ala Asp Ser Ile Arg Gln Gln Glu  
 1045 1050 1055  
 Leu Ser Ala Leu Arg Gln Asp Met Gln Glu Ala Gln Gly Glu Gln Lys  
 1060 1065 1070  
 Glu Leu Ser Ala Gln Met Glu Leu Leu Arg Gln Glu Val Lys Glu Lys  
 1075 1080 1085  
 Glu Ala Asp Phe Leu Ala Gln Glu Ala Gln Leu Leu Glu Glu Leu Glu  
 1090 1095 1100  
 Ala Ser His Ile Thr Glu Gln Gln Leu Arg Ala Ser Leu Trp Ala Gln  
 1105 1110 1115 1120

Glu Ala Lys Ala Ala Gln Leu Gln Leu Arg Leu Arg Ser Thr Glu Ser  
 1125 1130 1135  
 Gln Leu Glu Ala Leu Ala Ala Glu Gln Gln Pro Gly Asn Gln Ala Gln  
 1140 1145 1150  
 Ala Gln Ala Gln Leu Ala Ser Leu Tyr Ser Ala Leu Gln Gln Ala Leu  
 1155 1160 1165  
 Gly Ser Val Cys Glu Ser Arg Pro Glu Leu Ser Gly Gly Gly Asp Ser  
 1170 1175 1180  
 Ala Pro Ser Val Trp Gly Leu Glu Pro Asp Gln Asn Gly Ala Arg Ser  
 1185 1190 1195 1200  
 Leu Phe Lys Arg Gly Pro Leu Leu Thr Ala Leu Ser Ala Glu Ala Val  
 1205 1210 1215  
 Ala Ser Ala Leu His Lys Leu His Gln Asp Leu Trp Lys Thr Gln Gln  
 1220 1225 1230  
 Thr Arg Asp Val Leu Arg Asp Gln Val Gln Lys Leu Glu Glu Arg Leu  
 1235 1240 1245  
 Thr Asp Thr Glu Ala Glu Lys Ser Gln Val His Thr Glu Leu Gln Asp  
 1250 1255 1260  
 Leu Gln Arg Gln Leu Ser Gln Asn Gln Glu Glu Lys Ser Lys Trp Glu  
 1265 1270 1275 1280  
 Gly Lys Gln Asn Ser Leu Glu Ser Glu Leu Met Glu Leu His Glu Thr  
 1285 1290 1295  
 Met Ala Ser Leu Gln Ser Arg Leu Arg Arg Ala Glu Leu Gln Arg Met  
 1300 1305 1310  
 Glu Ala Gln Gly Glu Arg Glu Leu Leu Gln Ala Ala Lys Glu Asn Leu  
 1315 1320 1325  
 Thr Ala Gln Val Glu His Leu Gln Ala Ala Val Val Glu Ala Arg Ala  
 1330 1335 1340  
 Gln Ala Ser Ala Ala Gly Ile Leu Glu Glu Asp Leu Arg Thr Ala Arg  
 1345 1350 1355 1360  
 Ser Ala Leu Lys Leu Lys Asn Glu Glu Val Glu Ser Glu Arg Glu Arg  
 1365 1370 1375  
 Ala Gln Ala Leu Gln Glu Gln Gly Glu Leu Lys Val Ala Gln Gly Lys  
 1380 1385 1390  
 Ala Leu Gln Glu Asn Leu Ala Leu Leu Thr Gln Thr Leu Ala Glu Arg  
 1395 1400 1405  
 Glu Glu Glu Val Glu Thr Leu Arg Gly Gln Ile Gln Glu Leu Glu Lys  
 1410 1415 1420  
 Gln Arg Glu Met Gln Lys Ala Ala Leu Glu Leu Leu Ser Leu Asp Leu  
 1425 1430 1435 1440  
 Lys Lys Arg Asn Gln Glu Val Asp Leu Gln Gln Glu Gln Ile Gln Glu

1445	1450	1455
Leu Glu Lys Cys Arg Ser Val	Leu Glu His	Leu Pro Met Ala Val Gln
1460	1465	1470
Glu Arg Glu Gln Lys Leu Thr Val	Gln Arg Glu Gln Ile Arg Glu Leu	
1475	1480	1485
Glu Lys Asp Arg Glu Thr Gln Arg Asn Val	Leu Glu His Gln Leu Leu	
1490	1495	1500
Glu Leu Glu Lys Lys Asp Gln Met Ile Glu Ser Gln Arg Gly Gln Val		
1505	1510	1515 1520
Gln Asp Leu Lys Lys Gln Leu Val Thr Leu Glu Cys Leu Ala Leu Glu		
1525	1530	1535
Leu Glu Glu Asn His His Lys Met Glu Cys Gln Gln Lys Leu Ile Lys		
1540	1545	1550
Glu Leu Glu Gly Gln Arg Glu Thr Gln Arg Val Ala Leu Thr His Leu		
1555	1560	1565
Thr Leu Asp Leu Glu Glu Arg Ser Gln Glu Leu Gln Ala Gln Ser Ser		
1570	1575	1580
Gln Ile His Asp Leu Glu Ser His Ser Thr Val Leu Ala Arg Glu Leu		
1585	1590	1595 1600
Gln Glu Arg Asp Gln Glu Val Lys Ser Gln Arg Glu Gln Ile Glu Glu		
1605	1610	1615
Leu Gln Arg Gln Lys Glu His Leu Thr Gln Asp Leu Glu Arg Arg Asp		
1620	1625	1630
Gln Glu Leu Met Leu Gln Lys Glu Arg Ile Gln Val Leu Glu Asp Gln		
1635	1640	1645
Arg Thr Arg Gln Thr Lys Ile Leu Glu Glu Asp Leu Glu Gln Ile Lys		
1650	1655	1660
Leu Ser Leu Arg Glu Arg Gly Arg Glu Leu Thr Thr Gln Arg Gln Leu		
1665	1670	1675 1680
Met Gln Glu Arg Ala Glu Glu Gly Lys Gly Pro Ser Lys Ala Gln Arg		
1685	1690	1695
Gly Ser Leu Glu His Met Lys Leu Ile Leu Arg Asp Lys Glu Lys Glu		
1700	1705	1710
Val Glu Cys Gln Gln Glu His Ile His Glu Leu Gln Glu Leu Lys Asp		
1715	1720	1725
Gln Leu Glu Gln Gln Leu Gln Gly Leu His Arg Lys Val Gly Glu Thr		
1730	1735	1740
Ser Leu Leu Leu Ser Gln Arg Glu Gln Glu Ile Val Val Leu Gln Gln		
1745	1750	1755 1760
Gln Leu Gln Glu Ala Arg Glu Gln Gly Glu Leu Lys Glu Gln Ser Leu		
1765	1770	1775

Gln Ser Gln Leu Asp Glu Ala Gln Arg Ala Leu Ala Gln Arg Asp Gln  
 1780 1785 1790  
 Glu Leu Glu Ala Leu Gln Gln Glu Gln Gln Ala Gln Gly Gln Glu  
 1795 1800 1805  
 Glu Arg Val Lys Glu Lys Ala Asp Ala Leu Gln Gly Ala Leu Glu Gln  
 1810 1815 1820  
 Ala His Met Thr Leu Lys Glu Arg His Gly Glu Leu Gln Asp His Lys  
 1825 1830 1835 1840  
 Glu Gln Ala Arg Arg Leu Glu Glu Glu Leu Ala Val Glu Gly Arg Arg  
 1845 1850 1855  
 Val Gln Ala Leu Glu Glu Val Leu Gly Asp Leu Arg Ala Glu Ser Arg  
 1860 1865 1870  
 Glu Gln Glu Lys Ala Leu Leu Ala Leu Gln Gln Gln Cys Ala Glu Gln  
 1875 1880 1885  
 Ala Gln Glu His Glu Val Glu Thr Arg Ala Leu Gln Asp Ser Trp Leu  
 1890 1895 1900  
 Gln Ala Gln Ala Val Leu Lys Glu Arg Asp Gln Glu Leu Glu Ala Leu  
 1905 1910 1915 1920  
 Arg Ala Glu Ser Gln Ser Ser Arg His Gln Glu Glu Ala Ala Arg Ala  
 1925 1930 1935  
 Arg Ala Glu Ala Leu Gln Glu Ala Leu Gly Lys Ala His Ala Ala Leu  
 1940 1945 1950  
 Gln Gly Lys Glu Gln His Leu Leu Glu Gln Ala Glu Leu Ser Arg Ser  
 1955 1960 1965  
 Leu Glu Ala Ser Thr Ala Thr Leu Gln Ala Ser Leu Asp Ala Cys Gln  
 1970 1975 1980  
 Ala His Ser Arg Gln Leu Glu Glu Ala Leu Arg Ile Gln Glu Gly Glu  
 1985 1990 1995 2000  
 Ile Gln Asp Gln Asp Leu Arg Tyr Gln Glu Asp Val Gln Gln Leu Gln  
 2005 2010 2015  
 Gln Ala Leu Ala Gln Arg Asp Glu Glu Leu Arg His Gln Gln Glu Arg  
 2020 2025 2030  
 Glu Gln Leu Leu Glu Lys Ser Leu Ala Gln Arg Val Gln Glu Asn Met  
 2035 2040 2045  
 Ile Gln Glu Lys Gln Asn Leu Gly Gln Glu Arg Glu Glu Glu Glu Ile  
 2050 2055 2060  
 Arg Gly Leu His Gln Ser Val Arg Glu Leu Gln Leu Thr Leu Ala Gln  
 2065 2070 2075 2080  
 Lys Glu Gln Glu Ile Leu Glu Leu Arg Glu Thr Gln Gln Arg Asn Asn  
 2085 2090 2095

Leu Glu Ala Leu Pro His Ser His Lys Thr Ser Pro Met Glu Glu Gln  
 2100 2105 2110  
 Ser Leu Lys Leu Asp Ser Leu Glu Pro Arg Leu Gln Arg Glu Leu Glu  
 2115 2120 2125  
 Arg Leu Gln Ala Ala Leu Arg Gln Thr Glu Ala Arg Glu Ile Glu Trp  
 2130 2135 2140  
 Arg Glu Lys Ala Gln Asp Leu Ala Leu Ser Leu Ala Gln Thr Lys Ala  
 2145 2150 2155 2160  
 Ser Val Ser Ser Leu Gln Glu Val Ala Met Phe Leu Gln Ala Ser Val  
 2165 2170 2175  
 Leu Glu Arg Asp Ser Glu Gln Gln Arg Leu Gln Asp Glu Leu Glu Leu  
 2180 2185 2190  
 Thr Arg Arg Ala Leu Glu Lys Glu Arg Leu His Ser Pro Gly Ala Thr  
 2195 2200 2205  
 Ser Thr Ala Glu Leu Gly Ser Arg Gly Glu Gln Gly Val Gln Leu Gly  
 2210 2215 2220  
 Glu Val Ser Gly Val Glu Ala Glu Pro Ser Pro Asp Gly Met Glu Lys  
 2225 2230 2235 2240  
 Gln Ser Trp Arg Gln Arg Leu Glu His Leu Gln Gln Ala Val Ala Arg  
 2245 2250 2255  
 Leu Glu Ile Asp Arg Ser Arg Leu Gln Arg His Asn Val Gln Leu Arg  
 2260 2265 2270  
 Ser Thr Leu Glu Gln Asp Gly Arg Gly Gln Lys Asn Ser Asp Ala Lys  
 2275 2280 2285  
 Cys Val Ala Glu Leu Gln Lys Glu Val Val Leu Leu Gln Ala Gln Leu  
 2290 2295 2300  
 Thr Leu Glu Arg Lys Gln Lys Gln Asp Tyr Ile Thr Arg Ser Ala Gln  
 2305 2310 2315 2320  
 Thr Ser Arg Glu Leu Ala Gly Leu His His Ser Leu Ser His Ser Leu  
 2325 2330 2335  
 Leu Ala Val Ala Gln Ala Pro Glu Ala Thr Val Leu Glu Ala Glu Thr  
 2340 2345 2350  
 Arg Arg Leu Asp Glu Ser Leu Thr Gln Ser Leu Thr Ser Pro Gly Pro  
 2355 2360 2365  
 Val Leu Leu His Pro Ser Pro Ser Thr Thr Gln Ala Ala Ser Arg  
 2370 2375 2380

<210> 261  
 <211> 43  
 <212> PRT  
 <213> Homo sapiens  
 <400> 261

Met Tyr Arg Leu Ile Leu Phe Arg Asn Asn Ser Val Leu Glu Phe Ile  
 1 5 10 15

Lys Asn Ser Val Ile Ala Phe Ile Pro Lys Cys Leu Thr Leu Pro Thr  
 20 25 30

Ala Ser His Lys Ser Ile Tyr Phe Lys Ala Phe  
 35 40

<210> 262  
 <211> 34  
 <212> PRT  
 <213> Homo sapiens

<400> 262  
 Met Asp Pro Asn Phe Asp Ile Val His Thr Val Phe Ile Leu Cys Met  
 1 5 10 15

Glu Leu Ile Thr Asp Phe Ala Cys Lys Glu Arg Ile Val Cys Leu Asn  
 20 25 30

Phe Val

<210> 263  
 <211> 78  
 <212> PRT  
 <213> Homo sapiens

<400> 263  
 Met Met Glu Asn Ser Ala Pro Asn Ser Leu Met Asn Lys Glu Met Asp  
 1 5 10 15

His Leu Met Asp Glu Gly Val Gln Arg Thr Arg Val Ala Leu Gly Gln  
 20 25 30

Trp Leu Val Ala Ala Val Ile Gln Asp Leu Gly Ser Val Leu Cys Pro  
 35 40 45

Leu Pro Pro Ser Val Leu Ala Ser Arg Trp Gln Gly Val Ser Phe Pro  
 50 55 60

Glu Ser His Gln Leu Arg Gln Asn Pro Glu Ala Gly Lys Thr  
 65 70 75

<210> 264  
 <211> 85  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> UNSURE  
 <222> (15)..(72)

<400> 264  
 Met Gly Ile Tyr Ile Ile Tyr Ser Pro Arg Thr Val Ile Arg Xaa Xaa  
 1 5 10 15



Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
                   20                  25                  30  
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
                   35                  40                  45  
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
                   50                  55                  60  
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Leu Leu Ala Ser Phe Cys Phe Pro  
                   65                  70                  75                  80  
 Leu Val Leu Gly Phe  
                                   85

&lt;210&gt; 265

&lt;211&gt; 471

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 265

Leu Ser Phe Gln Ser Gly Asn Ile Ile Val Ala Thr Pro Gly Arg Leu  
   1                  5                  10                  15  
 Glu Asp Met Phe Arg Arg Lys Ala Glu Gly Leu Asp Leu Ala Ser Cys  
                   20                  25                  30  
 Val Arg Ser Leu Asp Val Leu Val Leu Asp Glu Ala Asp Arg Leu Leu  
                   35                  40                  45  
 Asp Met Gly Phe Glu Ala Ser Ile Asn Thr Ile Leu Glu Phe Leu Pro  
                   50                  55                  60  
 Lys Gln Arg Arg Thr Gly Leu Phe Ser Ala Thr Gln Thr Gln Glu Val  
                   65                  70                  75                  80  
 Glu Asn Leu Val Arg Ala Gly Leu Arg Asn Pro Val Arg Val Ser Val  
                   85                  90                  95  
 Lys Glu Lys Gly Val Ala Ala Ser Ser Ala Gln Lys Thr Pro Ser Arg  
                   100                  105                  110  
 Leu Glu Asn Tyr Tyr Met Val Cys Lys Ala Asp Glu Lys Phe Asn Gln  
                   115                  120                  125  
 Leu Val His Phe Leu Arg Asn His Lys Gln Glu Lys His Leu Val Phe  
                   130                  135                  140  
 Phe Gly Thr Cys Ala Cys Val Glu Tyr Tyr Gly Lys Ala Leu Glu Val  
                   145                  150                  155                  160  
 Leu Val Lys Gly Val Lys Ile Met Cys Ile His Gly Lys Met Lys Tyr  
                   165                  170                  175  
 Lys Arg Asn Lys Ile Phe Met Glu Phe Arg Lys Leu Gln Gly Gly Ile  
                   180                  185                  190  
 Leu Val Cys Thr Asp Val Met Ala Arg Gly Ile Asp Ile Pro Glu Val  
                   195                  200                  205

Asn Trp Val Leu Gln Tyr Asp Pro Pro Ser Asn Ala Ser Ala Phe Val  
 210 215 220  
 His Arg Cys Gly Arg Thr Ala Arg Ile Gly His Gly Gly Ser Ala Leu  
 225 230 235 240  
 Val Phe Leu Leu Pro Met Glu Glu Ser Tyr Ile Asn Phe Leu Ala Ile  
 245 250 255  
 Asn Gln Lys Cys Pro Leu Gln Glu Met Lys Pro Gln Arg Asn Thr Ala  
 260 265 270  
 Asp Leu Leu Pro Lys Leu Lys Ser Met Ala Leu Ala Asp Arg Ala Val  
 275 280 285  
 Phe Glu Lys Gly Met Lys Ala Phe Val Ser Tyr Val Gln Ala Tyr Ala  
 290 295 300  
 Lys His Glu Cys Asn Leu Ile Phe Arg Leu Lys Asp Leu Asp Phe Ala  
 305 310 315 320  
 Ser Leu Ala Arg Gly Phe Ala Leu Leu Arg Met Pro Lys Met Pro Glu  
 325 330 335  
 Leu Arg Gly Lys Gln Phe Pro Asp Phe Val Pro Val Asp Val Asn Thr  
 340 345 350  
 Asp Thr Ile Pro Phe Lys Asp Lys Ile Arg Glu Lys Gln Arg Gln Lys  
 355 360 365  
 Leu Leu Glu Gln Gln Arg Arg Glu Lys Thr Glu Asn Glu Gly Arg Arg  
 370 375 380  
 Lys Phe Ile Lys Asn Lys Ala Trp Ser Lys Gln Lys Ala Lys Lys Glu  
 385 390 395 400  
 Lys Lys Lys Lys Met Asn Glu Lys Arg Lys Arg Glu Glu Gly Ser Asp  
 405 410 415  
 Ile Glu Asp Glu Asp Met Glu Glu Leu Leu Asn Asp Thr Arg Leu Leu  
 420 425 430  
 Lys Lys Leu Lys Lys Gly Lys Ile Thr Glu Glu Glu Phe Glu Lys Gly  
 435 440 445  
 Leu Leu Thr Thr Gly Lys Arg Thr Ile Lys Thr Val Asp Leu Gly Ile  
 450 455 460  
 Ser Asp Leu Glu Asp Asp Cys  
 465 470

&lt;210&gt; 266

&lt;211&gt; 20

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 266

Met Met Thr Ser Leu Ser Tyr Ser Ser Gln Ser Trp Lys Pro Cys Ser  
 1 5 10 15

Gln Ser Phe Lys  
20

<210> 267  
<211> 27  
<212> PRT  
<213> Homo sapiens

<400> 267  
Met Val Phe Leu Glu Ile Ile Phe Cys Pro Met Tyr Ser Ile Phe Ile  
1 5 10 15  
His Thr Gly Phe Ile Met Ile Ile Ile Ser Lys  
20 25

<210> 268  
<211> 55  
<212> PRT  
<213> Homo sapiens

<400> 268  
Met Leu Arg Gly Asp Leu Pro Gly Ser Val Leu Pro Leu Ser Leu Arg  
1 5 10 15  
Leu Asn Gly Ala Pro Pro Arg Leu Leu Pro Gly Lys Lys His Ser Gly  
20 25 30  
Gln Ala Gly Pro Glu Pro Val Ser Val Arg Gly Pro Val Ala Cys Pro  
35 40 45  
Gly Gly Arg Ser Leu Gln Gly  
50 55

<210> 269  
<211> 38  
<212> PRT  
<213> Homo sapiens

<400> 269  
Met Asn Glu Ala Asn Arg Leu Phe Phe Val Ser Leu Thr Pro Arg Asn  
1 5 10 15  
Ile Met Ile Pro Tyr Lys Ile Leu Ile His Thr His Asp Gln Tyr Phe  
20 25 30  
Ile Pro Thr Glu Thr Val  
35

<210> 270  
<211> 71  
<212> PRT  
<213> Homo sapiens

<400> 270  
Met Leu Thr Leu Val Tyr Leu Val Val Glu Asn Gly Leu Leu Pro Leu  
1 5 10 15

Phe Pro Glu Leu Thr Leu Phe Pro Leu Ala Arg Arg Ser Gly Gln Arg  
                     20                    25                    30  
 Glu Pro Arg Thr Glu Val Pro Thr Thr Gln Gln Ala Leu Ser Ser Pro  
                     35                    40                    45  
 Leu Thr Ser Asn Val Cys Ile His Phe Gln Pro Leu Thr Asp Leu Val  
                     50                    55                    60  
 Phe Gln Cys Ile Ile Ile Leu  
                     65                    70

<210> 271  
 <211> 65  
 <212> PRT  
 <213> Homo sapiens

<400> 271  
 Met Glu Glu Ser Lys Ala Gln Arg Arg Arg Glu Thr Thr Trp Ser Val  
                     1                    5                    10                    15  
 Ser Leu Ser Gln Leu Ile Gln His Pro Thr Asn His Pro Ser His Ser  
                     20                    25                    30  
 Leu Ser Ile Ser Leu Val Asn Trp Ser Thr Ile Cys Asn Cys Ser Gln  
                     35                    40                    45  
 Val Pro Pro Asn Ser Leu Cys Arg Tyr Phe Ser Cys Val Phe His Ser  
                     50                    55                    60  
 Leu  
                     65

<210> 272  
 <211> 25  
 <212> PRT  
 <213> Homo sapiens

<400> 272  
 Met Val Pro Ile Ile Ser Tyr Val Lys Met Ser Cys Tyr Glu Lys Leu  
                     1                    5                    10                    15  
 Phe Leu Phe Gln Ser Cys Gln Cys Gln  
                     20                    25

<210> 273  
 <211> 13  
 <212> PRT  
 <213> Homo sapiens

<400> 273  
 Met Leu Leu Ser Tyr Ser Ala Gln Glu Tyr Leu Ser Lys  
                     1                    5                    10

<210> 274  
 <211> 73  
 <212> PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 274

Met Lys Cys Val Ser Glu His Gln Arg Pro Ser Ile Leu Pro Leu Pro  
 1 5 10 15

Phe Leu Val Val Tyr Lys Asn Ser Arg Leu Glu Glu Phe Arg Phe Val  
 20 25 30

Ala His Phe Phe Pro Gln His Phe Phe Leu Leu Phe Phe Lys Met Tyr  
 35 40 45

Cys Leu Phe Pro His Ser Val Thr Leu Asp Ile Gly Ile Phe Asn Cys  
 50 55 60

Val Ile Phe Cys Cys Lys Lys Gly Lys  
 65 70

&lt;210&gt; 275

&lt;211&gt; 465

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 275

Met Leu Gly Ser Met Ala Arg Lys Lys Pro Arg Asn Thr Ser Arg Leu  
 1 5 10 15

Pro Leu Ala Leu Asn Pro Leu Lys Ser Lys Asp Val Leu Ala Val Leu  
 20 25 30

Ala Glu Arg Asn Glu Ala Ile Val Pro Val Gly Ala Trp Val Glu Pro  
 35 40 45

Ala Ser Pro Gly Ser Ser Glu Ile Pro Ala Tyr Thr Ser Ala Tyr Leu  
 50 55 60

Ile Glu Glu Glu Leu Lys Glu Gln Leu Arg Lys Lys Gln Glu Ala Leu  
 65 70 75 80

Lys His Phe Gln Lys Gln Val Lys Tyr Arg Val Asn Gln Gln Ile Arg  
 85 90 95

Leu Arg Lys Lys Gln Gln Leu Gln Lys Ser Tyr Glu Arg Ala Gln Lys  
 100 105 110

Glu Gly Ser Ile Ala Met Gln Ser Ser Ala Thr His Leu Thr Ser Lys  
 115 120 125

Arg Thr Ser Val Phe Pro Asn Asn Leu Asn Val Ala Ile Gly Ser Ser  
 130 135 140

Arg Leu Pro Pro Ser Leu Met Pro Gly Asp Gly Ile Glu Asp Glu Glu  
 145 150 155 160

Asn Gln Asn Glu Leu Phe Gln Gln Gln Ala Gln Ala Leu Ser Glu Thr  
 165 170 175

Met Lys Gln Ala Arg His Arg Leu Ala Ser Phe Lys Thr Val Ile Lys  
 180 185 190

Lys Lys Gly Ser Val Phe Pro Asp Asp Gly Arg Lys Ser Phe Leu Thr  
 195 200 205  
 Arg Glu Glu Val Leu Ser Arg Lys Pro Ala Ser Thr Gly Ile Asn Thr  
 210 215 220  
 Gly Ile Arg Gly Glu Leu Pro Ile Lys Val His Gln Gly Leu Leu Ala  
 225 230 235 240  
 Ala Val Pro Tyr Gln Asn Tyr Met Glu Asn Gln Glu Leu Asp Tyr Glu  
 245 250 255  
 Glu Pro Asp Tyr Glu Glu Ser Ser Ser Leu Val Thr Asp Glu Lys Gly  
 260 265 270  
 Lys Glu Asp Leu Phe Gly Arg Gly Gln Gln Asp Gln Gln Ala Ile His  
 275 280 285  
 Ser Glu Asp Lys Asn Lys Pro Phe Ser Arg Val Gln Lys Val Lys Phe  
 290 295 300  
 Lys Asn Pro Leu Phe Val Leu Met Glu Glu Glu Glu Gln Lys Gln Leu  
 305 310 315 320  
 His Phe Glu Gly Leu Gln Asp Ile Leu Pro Glu Ala Gln Asp Tyr Phe  
 325 330 335  
 Leu Glu Ala Gln Gly Asp Leu Leu Glu Thr Gln Gly Asp Leu Thr Gly  
 340 345 350  
 Ile Gln Ser Val Lys Pro Asp Thr Gln Ala Val Glu Met Lys Val Gln  
 355 360 365  
 Val Thr Glu Pro Glu Gly Gln Ala Ile Glu Pro Glu Gly Gln Pro Ile  
 370 375 380  
 Lys Thr Glu Thr Gln Gly Ile Met Leu Lys Ala Gln Ser Ile Glu Leu  
 385 390 395 400  
 Glu Glu Gly Ser Ile Val Leu Lys Thr Gln Asp Phe Leu Pro Thr Asn  
 405 410 415  
 Gln Ala Leu Leu Thr Lys Asn Gln Asp Val Leu Leu Lys Asp His Cys  
 420 425 430  
 Val Leu Pro Lys Asp Gln Ser Ile Leu Leu Lys Tyr Gln Asp Gln Asp  
 435 440 445  
 Phe Leu Pro Arg Asp Gln His Val Leu His Lys Asp Gln Asp Ile Leu  
 450 455 460

Pro  
 465

<210> 276

<211> 38

<212> PRT

<213> Homo sapiens

<400> 276

Met Asn Lys Gln Lys Ile Lys Met Phe Arg Met Lys Ile Leu Leu Lys  
 1 5 10 15  
 Trp Ser Leu Glu Ile Thr Val Met Ser Ala Leu Gly Ile Glu Ser Arg  
 20 25 30  
 Ile Asn Ser Gln Ile Pro  
 35

<210> 277  
 <211> 170  
 <212> PRT  
 <213> Homo sapiens

<400> 277  
 Met Asp Ile Glu Arg Glu Gln Val Lys Glu Gln Gln Arg Gln Lys Glu  
 1 5 10 15  
 Gln Lys Lys Lys Ile Glu Lys Ile Lys Lys Lys Arg Glu Gln Glu Cys  
 20 25 30  
 Tyr Ala Ala Glu Gln Arg Ile Leu Arg Met Asn Phe His Glu Asp Pro  
 35 40 45  
 Tyr Ser Gly Glu Lys Leu Ser Glu Ile Leu Ala Gln Leu Gln Leu Gln  
 50 55 60  
 Glu Ile Lys Gly Thr Arg Glu Lys Gln Gln Arg Glu Lys Glu Tyr Leu  
 65 70 75 80  
 Arg Tyr Val Glu Ala Leu Arg Ala Gln Ile Gln Glu Lys Met Gln Leu  
 85 90 95  
 Tyr Asn Ile Thr Leu Pro Pro Leu Cys Cys Cys Gly Pro Asp Phe Trp  
 100 105 110  
 Asp Ala His Pro Asp Thr Cys Ala Asn Asn Cys Ile Phe Tyr Lys Asn  
 115 120 125  
 His Arg Ala Tyr Thr Arg Ala Leu His Ser Phe Ile Asn Ser Cys Asp  
 130 135 140  
 Val Pro Gly Gly Asn Ser Thr Leu Arg Val Ala Ile His Asn Phe Ala  
 145 150 155 160  
 Ser Ala His Arg Arg Thr Leu Lys Asn Leu  
 165 170

<210> 278  
 <211> 173  
 <212> PRT  
 <213> Homo sapiens

<400> 278  
 Ala Tyr Asp Arg Tyr Gln Ser Gly Leu Ser Thr Glu Phe Gln Ala Pro  
 1 5 10 15  
 Leu Ala Phe Gln Ser Asp Val Asp Lys Glu Glu Asp Lys Lys Glu Arg  
 20 25 30

Gln Lys Gln Tyr Leu Arg His Arg Arg Leu Phe Met Asp Ile Glu Arg  
           35                          40                          45  
 Glu Gln Val Lys Glu Gln Gln Arg Gln Lys Glu Gln Lys Lys Lys Ile  
           50                          55                          60  
 Glu Lys Ile Lys Lys Lys Arg Glu Gln Glu Cys Tyr Ala Ala Glu Gln  
           65                          70                          75                          80  
 Arg Ile Leu Arg Met Asn Phe His Glu Asp Pro Tyr Ser Gly Glu Lys  
                           85                          90                          95  
 Leu Ser Glu Ile Leu Ala Gln Leu Gln Leu Gln Glu Ile Lys Gly Thr  
                           100                          105                          110  
 Arg Glu Lys Gln Gln Arg Glu Lys Glu Tyr Leu Arg Tyr Val Glu Ala  
                           115                          120                          125  
 Leu Arg Ala Gln Ile Gln Glu Lys Met Gln Leu Tyr Asn Ile Thr Leu  
                           130                          135                          140  
 Pro Pro Leu Cys Cys Cys Gly Pro Asp Phe Trp Asp Ala His Pro Asp  
           145                          150                          155                          160  
 Thr Cys Ala Asn Asn Cys Ile Phe Tyr Lys Asn His Arg  
                           165                          170

<210> 279  
 <211> 15  
 <212> PRT  
 <213> Homo sapiens

<400> 279  
 Met Ile Ser Arg Ile Leu Pro Phe Ile Tyr Ser Thr Ser Ile Arg  
       1                          5                          10                          15

<210> 280  
 <211> 11  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> UNSURE  
 <222> (8)

<400> 280  
 Met Asp Thr Gly Leu Phe Phe Xaa Gly Ala Gly  
       1                          5                          10

<210> 281  
 <211> 86  
 <212> PRT  
 <213> Homo sapiens

<400> 281  
 Met Ala Val Ser Leu Phe Leu Ser Ala Asp Pro Ser Met Thr Leu Ile  
       1                          5                          10                          15



Arg Phe Pro Phe Ser Tyr Asn Ser Cys Pro Trp Ile Gln Trp Pro Ser  
                   20                  25                  30  
 Phe Phe Ser Phe Ala Leu Phe Ser Val Thr Val His His Ile Phe Tyr  
                   35                  40                  45  
 Thr Ala Val Asp Val Ile Tyr Ser Asn Asp Val Pro Val Pro Phe Val  
           50                  55                  60  
 Cys Leu Phe Leu Glu Thr Pro Ser Gly Ala Phe His Leu Pro Gly Ser  
       65                  70                  75                  80  
 Asn Leu Asp Trp Leu Leu  
                   85

<210> 282  
 <211> 1339  
 <212> PRT  
 <213> Homo sapiens

<400> 282  
 Met Ala Val Tyr Cys Tyr Ala Leu Asn Ser Leu Val Ile Met Asn Ser  
       1                  5                  10                  15  
 Ala Asn Glu Met Lys Ser Gly Gly Gly Pro Gly Pro Ser Gly Ser Glu  
                   20                  25                  30  
 Thr Pro Pro Pro Pro Arg Arg Ala Val Leu Ser Pro Gly Ser Val Phe  
           35                  40                  45  
 Ser Pro Gly Arg Gly Ala Ser Phe Leu Phe Pro Pro Ala Glu Ser Leu  
       50                  55                  60  
 Ser Pro Glu Glu Pro Arg Ser Pro Gly Gly Trp Arg Ser Gly Arg Arg  
       65                  70                  75                  80  
 Arg Leu Asn Ser Ser Ser Gly Ser Gly Ser Gly Ser Ser Gly Ser Ser  
                   85                  90                  95  
 Val Ser Ser Pro Ser Trp Ala Gly Arg Leu Arg Gly Asp Arg Gln Gln  
           100                  105                  110  
 Val Val Ala Ala Gly Thr Leu Ser Pro Pro Gly Pro Glu Glu Ala Lys  
           115                  120                  125  
 Arg Lys Leu Arg Ile Leu Gln Arg Glu Leu Gln Asn Val Gln Val Asn  
       130                  135                  140  
 Gln Lys Val Gly Met Phe Glu Ala His Ile Gln Ala Gln Ser Ser Ala  
       145                  150                  155                  160  
 Ile Gln Ala Pro Arg Ser Pro Arg Leu Gly Arg Ala Arg Ser Pro Ser  
           165                  170                  175  
 Pro Cys Pro Phe Arg Ser Ser Ser Gln Pro Pro Gly Arg Val Leu Val  
           180                  185                  190  
 Gln Gly Ala Arg Ser Glu Glu Arg Arg Thr Lys Ser Trp Gly Glu Gln  
       195                  200                  205

Cys Pro Glu Thr Ser Gly Thr Asp Ser Gly Arg Lys Gly Gly Pro Ser  
 210 215 220  
 Leu Cys Ser Ser Gln Val Lys Lys Gly Met Pro Pro Leu Pro Gly Arg  
 225 230 235 240  
 Ala Ala Pro Thr Gly Ser Glu Ala Gln Gly Pro Ser Ala Phe Val Arg  
 245 250 255  
 Met Glu Lys Gly Ile Pro Ala Ser Pro Arg Cys Gly Ser Pro Thr Ala  
 260 265 270  
 Met Glu Ile Asp Lys Arg Gly Ser Pro Thr Pro Gly Thr Arg Ser Cys  
 275 280 285  
 Leu Ala Pro Ser Leu Gly Leu Phe Gly Ala Ser Leu Thr Met Ala Thr  
 290 295 300  
 Glu Val Ala Ala Arg Val Thr Ser Thr Gly Pro His Arg Pro Gln Asp  
 305 310 315 320  
 Leu Ala Leu Thr Glu Pro Ser Gly Arg Ala Arg Glu Leu Glu Asp Leu  
 325 330 335  
 Gln Pro Pro Glu Ala Leu Val Glu Arg Gln Gly Gln Phe Leu Gly Ser  
 340 345 350  
 Glu Thr Ser Pro Ala Pro Glu Arg Gly Gly Pro Arg Asp Gly Glu Pro  
 355 360 365  
 Pro Gly Lys Met Gly Lys Gly Tyr Leu Pro Cys Gly Met Pro Gly Ser  
 370 375 380  
 Gly Glu Pro Glu Val Gly Lys Arg Pro Glu Glu Thr Thr Val Ser Val  
 385 390 395 400  
 Gln Ser Ala Glu Ser Ser Asp Ser Leu Ser Trp Ser Arg Leu Pro Arg  
 405 410 415  
 Ala Leu Ala Ser Val Gly Pro Glu Glu Ala Arg Ser Gly Ala Pro Val  
 420 425 430  
 Gly Gly Gly Arg Trp Gln Leu Ser Asp Arg Val Glu Gly Gly Ser Pro  
 435 440 445  
 Thr Leu Gly Leu Leu Gly Gly Ser Pro Ser Ala Gln Pro Gly Thr Gly  
 450 455 460  
 Asn Val Glu Ala Gly Ile Pro Ser Gly Arg Met Leu Glu Pro Leu Pro  
 465 470 475 480  
 Cys Trp Asp Ala Ala Lys Asp Leu Lys Glu Pro Gln Cys Pro Pro Gly  
 485 490 495  
 Asp Arg Val Gly Val Gln Pro Gly Asn Ser Arg Val Trp Gln Gly Thr  
 500 505 510  
 Met Glu Lys Ala Gly Leu Ala Trp Thr Arg Gly Thr Gly Val Gln Ser  
 515 520 525

Glu Gly Thr Trp Glu Ser Gln Arg Gln Asp Ser Asp Ala Leu Pro Ser  
 530 535 540  
 Pro Glu Leu Leu Pro Gln Asp Pro Asp Lys Pro Phe Leu Arg Lys Ala  
 545 550 555 560  
 Cys Ser Pro Ser Asn Ile Pro Ala Val Ile Ile Thr Asp Met Gly Thr  
 565 570 575  
 Gln Glu Asp Gly Ala Leu Glu Glu Thr Gln Gly Ser Pro Arg Gly Asn  
 580 585 590  
 Leu Pro Leu Arg Lys Leu Ser Ser Ser Ser Ala Ser Ser Thr Gly Phe  
 595 600 605  
 Ser Ser Ser Tyr Glu Asp Ser Glu Glu Asp Ile Ser Ser Asp Pro Glu  
 610 615 620  
 Arg Thr Leu Asp Pro Asn Ser Ala Phe Leu His Thr Leu Asp Gln Gln  
 625 630 635 640  
 Lys Pro Arg Val Lys Tyr Arg Thr Ile Trp Lys Val Lys Asn Lys Glu  
 645 650 655  
 Arg Glu Ser Ser Pro Gly Asn Ala Ser Leu Leu Leu Ile Pro Val Thr  
 660 665 670  
 Ala Ala Thr Gly Ile Arg Val Leu Gly Leu Gly Leu Gly Asp Leu Gly  
 675 680 685  
 Glu Ile Pro Val Tyr Thr Trp Leu Ala Ser Ser Leu Lys Asn Gly Glu  
 690 695 700  
 Ser Lys Cys Asp Leu Met Glu Trp Tyr Cys Tyr Thr Val Lys His Pro  
 705 710 715 720  
 Gly Ser Leu Glu Leu His Gly Leu Arg Met Ser Pro Thr Gly Thr Ser  
 725 730 735  
 Cys Cys Gly Leu Ile Met Ser Ala Pro Lys Gln Glu Leu Asn Ala Ile  
 740 745 750  
 Glu Leu Ser Tyr Leu Pro Pro Ala Pro Ile Val Val Val Arg Lys Ser  
 755 760 765  
 Gly Phe Ser Ala Gln Gln Ser Ala Trp Asp Cys Ile Lys Pro Ser Ser  
 770 775 780  
 Pro Ile Arg Asp Arg Val Ala Leu Leu Cys Pro Met Gly Phe Lys Ala  
 785 790 795 800  
 Lys Gly Leu Tyr Glu Ser Cys Leu Trp His Ser Pro Glu Ser Ser Gly  
 805 810 815  
 Ile Arg Gln Lys Gln Cys Cys Ala Ala Leu Ser Trp Ala Leu Lys Gly  
 820 825 830  
 Lys Arg Glu Tyr Leu Gln Gln Tyr Ser Gly Trp Met Trp Val Pro Gly  
 835 840 845  
 Leu Leu Ile Leu Gly Leu Gly Leu Ser Glu Ile His Arg Ser Ser Leu

850	855	860
Gln Val Gln Pro Ala Gly	Gly Val His Thr Glu	Ala Ala Ala Pro Gly
865	870	875 880
Ala Pro Gly His Gln Gly	Ala Met Ser Val Thr Tyr	Asp Ala Leu Arg
885	890	895
Glu Lys Gln Gln Leu Ser	Lys Val Gly Asp Leu Pro	Ala Leu Thr Trp
900	905	910
Pro Gly Pro Leu Ile Ser	Gln Met Pro Gly Val Leu	Asp Ser Cys Arg
915	920	925
Leu Cys Ser Leu Gly Asp	Ile Glu Lys Ser Lys Ser	Trp Arg Lys Ile
930	935	940
Lys Asn Met Val His Trp	Ser Pro Phe Val Met Ser	Phe Lys Lys Lys
945	950	955 960
Tyr Pro Trp Ile Gln Leu	Ala Gly His Ala Gly Ser	Phe Lys Ala Ala
965	970	975
Ala Asn Gly Arg Ile Leu	Lys Lys His Cys Glu Ser	Glu Gln Arg Cys
980	985	990
Leu Asp Arg Leu Met Val	Asp Val Leu Arg Pro Phe	Val Pro Ala Tyr
995	1000	1005
His Gly Asp Val Val Lys	Asp Gly Glu Arg Tyr Asn	Gln Met Asp Asp
1010	1015	1020
Leu Leu Ala Asp Phe Asp	Ser Pro Cys Val Met Asp	Cys Lys Met Gly
1025	1030	1035 1040
Ile Arg Gln Gln Gln Asp	Phe Ala Gly Asp His Met	Glu Asn Asn Pro
1045	1050	1055
Ser Gly Val His Ser Asp	Leu Ala Lys Lys Ala Gly	Glu Cys Gly Glu
1060	1065	1070
Gly Leu Ser Leu Thr Phe	Leu Trp Ala Ser Arg Pro	Thr Ile Gln Leu
1075	1080	1085
Ala Pro Pro Val Asp Ile	Ser Pro Gln Pro Leu Ser	Ser Pro Gly Gln
1090	1095	1100
Thr Tyr Leu Glu Glu Glu	Leu Thr Lys Ala Arg Lys	Lys Pro Ser Leu
1105	1110	1115 1120
Arg Lys Asp Met Tyr Gln	Lys Met Ile Glu Val Asp	Pro Glu Ala Pro
1125	1130	1135
Thr Glu Glu Glu Lys Ala	Gln Arg Ala Val Thr Lys	Pro Arg Tyr Met
1140	1145	1150
Gln Trp Arg Glu Thr Ile	Ser Ser Thr Ala Thr Leu	Gly Phe Arg Ile
1155	1160	1165
Glu Gly Ile Lys Leu Arg	Gly Ser Ala Trp Gly Ala	Leu Pro Thr Ala
1170	1175	1180

Pro Gly Ser Arg Pro Leu Leu His Pro Gly Leu Leu Pro Gln Pro Gln  
 1185 1190 1195 1200  
 Val Leu Pro Val Leu Ser Lys Ala Ala Thr Lys Glu Asp Gly Thr Val  
 1205 1210 1215  
 Asn Arg Asp Phe Lys Lys Thr Lys Thr Arg Glu Gln Val Thr Glu Ala  
 1220 1225 1230  
 Phe Arg Glu Phe Thr Lys Gly Asn His Asn Ile Leu Ile Ala Tyr Arg  
 1235 1240 1245  
 Asp Arg Leu Lys Ala Ile Arg Thr Thr Leu Glu'Val Ser Pro Phe Phe  
 1250 1255 1260  
 Lys Cys His Glu Val Ile Gly Ser Ser Leu Leu Phe Ile His Asp Lys  
 1265 1270 1275 1280  
 Lys Glu Gln Ala Lys Val Trp Met Ile Asp Phe Gly Lys Thr Thr Pro  
 1285 1290 1295  
 Leu Pro Glu Gly Gln Thr Leu Gln His Asp Val Pro Trp Gln Glu Gly  
 1300 1305 1310  
 Asn Arg Glu Asp Gly Tyr Leu Ser Gly Leu Asn Asn Leu Val Asp Ile  
 1315 1320 1325  
 Leu Thr Glu Met Ser Gln Asp Ala Pro Leu Ala  
 1330 1335